



**SOLID STATE
ELECTRONICS
LABORATORY**

①

AD-A261 948



2

JSEP ANNUAL REPORT

11 January, 1992 through 10 January, 1993

**James S. Harris, Jr.
JSEP Principal Investigator
and Program Director**

(415) 723-9775

**This work was supported by the
Joint Services Electronics Program
(U.S. Army, U.S. Navy and U.S. Air Force)
Contract DAAL03-91-C-0010,
and was monitored by the
U.S. Army Research Office**

**DTIC
ELECTE
MAR 29 1993
S E D**

**Reproduction in whole or in part is permitted
for any purpose of the United States Government**

**This document has been approved for public
release and sale; its distribution is unlimited**

**STANFORD ELECTRONICS LABORATORIES • DEPARTMENT OF ELECTRICAL ENGINEERING
STANFORD UNIVERSITY • STANFORD, CA 94305-4055**

93 3 25 066

469 505
93-06127



10208

8c. ADDRESS (City, State, and ZIP Code)

P. O. Box 12211
Research Triangle Park, NC 27709-2211

10. SOURCE OF FUNDING NUMBERS

PROGRAM
ELEMENT NO.PROJECT
NO.TASK
NO.WORK UNIT
ACCESSION NO.

11. TITLE (Include Security Classification)

JSEP Annual Report January 11, 1992 - January 10, 1993

12. PERSONAL AUTHOR(S)

J. S. Harris, Principal Investigator and Program Director

13a. TYPE OF REPORT
Annual13b. TIME COVERED
FROM 1/11/92 TO 1/10/9314. DATE OF REPORT (Year, Month, Day)
3/9315. PAGE COUNT
99

16. SUPPLEMENTARY NOTATION

The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

17. COSATI CODES

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)

| FIELD | GROUP | SUB-GROUP |
|-------|-------|-----------|
| | | |
| | | |
| | | |

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

See Attached

This is the annual report of the research conducted at the Stanford Electronics Laboratories under the sponsorship of the Joint Services Electronics Program from January 11, 1992 through January 10, 1993. This report summarizes the area of research, identifies the most significant results and lists the dissertations and publications sponsored by the contract DAAL03-91-C-0010.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT

☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION

Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL

J. S. Harris

22b. TELEPHONE (Include Area Code)

415-723-9775

22c. OFFICE SYMBOL

JSEP ANNUAL REPORT

11 January, 1992 through 10 January, 1993

James S. Harris, Jr.
JSEP Principal Investigator
and Program Director

(415) 723-9775

This work was supported by the
Joint Services Electronics Program
(U.S. Army, U.S. Navy and U.S. Air Force)
Contract DAAL03-91-C-0010,
and was monitored by the
U.S. Army Research Office

| | |
|---------------------|---|
| Accession For | |
| NTIS | CRA&I <input checked="" type="checkbox"/> |
| DTIC | TAB <input type="checkbox"/> |
| Unannounced | <input type="checkbox"/> |
| Justification | |
| By | |
| Distribution / | |
| Availability Codes | |
| Dist | Avail and/or Special |
| A-1 | |

Reproduction in whole or in part is permitted
for any purpose of the United States Government

This document has been approved for public
release and sale; its distribution is unlimited

DTIC QUALITY INSPECTED 1

Abstract

This is the annual report of the research conducted at the Stanford Electronics Laboratories under the sponsorship of the Joint Services Electronics Program from 11 January 1992 through 10 January 1993. This report summarizes the area of research, identifies the most significant results and lists the dissertations and publications sponsored by the contract DAAL03-91-C-0010.

Key Words and Phrases: None

TABLE OF CONTENTS

| | |
|--|-----------|
| Introduction | 2 |
| Unit 1: Investigation of Transport in Quantum Dots | 5 |
| Unit 2: Physics and Applications of Ultra-Small Structures | 15 |
| Unit 3: Reactive Ion Profiling of Heterostructures | 21 |
| Unit 4: The Electronic Structure and Interfacial Properties of High Temperature Superconductors | 31 |
| Unit 5: Semiconductor Laser Structures for Optical Interconnects | 39 |
| Unit 6: Quantum Computing | 45 |
| Unit 7: Applications of Silicon-Germanium and Germanium Films in MOS Technologies | 53 |
| Unit 8: Signal Processing for Wideband Digital Portable Communications | 63 |
| Unit 9: Wideband Data Transmission Techniques | 73 |
| Unit 10: Fast Computing with Neural Networks | 85 |
| Unit 11: Research in Image Compression, Data Compression and Network Information Flow | 93 |

This work was supported by the Joint Services Electronics Program, contract DAAL03-91-C-0010. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies either expressed or implied, of the U.S. Government.

JSEP ANNUAL REPORT

January 11, 1992 - January 10, 1993

Introduction

The JSEP contract supports a program of unclassified basic research in electronics conducted by faculty members of the Electrical Engineering Department of Stanford University as a component of the research program of the Stanford Electronics Laboratories. The Stanford Electronics Lab JSEP Director and Principal Investigator is Professor James Harris. He is responsible for the selection of the best individual proposals, coordination between Stanford and the JSEP TCC and coordination between the selected areas of the JSEP Program. In planning the JSEP Program at SEL, a general objective is to develop new projects with 3-4 years of JSEP sponsorship, leading to a transition to more conventional DoD or other agency program funding. Since this type of funding often requires 12-18 months in the proposal, evaluation and budgeting stages, the flexibility in JSEP funding allows us to seize new opportunities and initiate programs which might otherwise be delayed for a significant period. An outstanding example of this funding initiation was the Molecular Beam Epitaxial growth of high temperature superconductors four years ago, which was supported by JSEP for two years while the DoD high T_c programs were being organized, proposals evaluated and finally funded. This project was the nucleation for the DARPA supported collaborative program with Varian Associates. Similarly, the observation of strong visible light emission for porous Si stimulated a change in the Ultra-Small Structures work unit which has produced very impressive results in a short time that would have been impossible to achieve without the flexibility afforded by JSEP funding.

The technical knowledge developed under the JSEP contract is widely disseminated through sponsor reviews, presentations of papers at technical meetings, publications in the open literature, discussions with visitors to the laboratories, and publication of laboratory technical reports (Ph.D. dissertations). Highlights of four projects are described below. Following these highlights, the specific objectives and progress in each work unit are reported.

The recent observations of room temperature visible photoluminescence and electroluminescence from electrochemically

etched porous silicon has rekindled interest in the potential of using silicon as an optical material. We have successfully fabricated high aspect ratio Si nano-wires using a combination of high resolution electron beam lithography and anisotropic reactive ion etching. We demonstrated that it is possible to engineer the profile of the Si nano-wires by intentional introduction of Chromium (Cr) into the Chlorine (Cl₂) reactive ion etching chamber. With thermal oxidation, the Si core diameter can be further reduced to below 10 nm. A transmission electron microscopy (TEM) technique was developed to obtain the precise structural information of the Si nano-wires and lattice images of the Si cores surrounded by oxide skins were obtained with this technique. Photoluminescence was also observed from these Si nano-wires.

Low temperature processing is increasingly important, especially for large area liquid crystal displays which are on glass substrates. Si films become amorphous when deposited below 600°C, thus limiting the quality of low temperature films for thin film transistors (TFT) and poly gates for the MOSFETs required for the active matrix switches in these displays. Poly-SiGe films with Ge mole fractions up to 0.6 are completely compatible with standard VLSI fabrication processes. We demonstrated deposition in commercially available LPCVD equipment using silane (SiH₄) and germane (GeH₄) as the gaseous deposition sources at a temperature of 400°C. these films can also be crystallized at lower temperature (550°C), and annealed for dopant activation at lower temperature (500°C) than pure poly-Si films. We also demonstrated that grain boundaries in such films can be passivated by hydrogen implantation.

Nd_{2-x}Ce_xCuO_{4+d} (NCCO) and its relatives are particularly interesting because they are the only n-type HTSCs that have been discovered so far. All of the n-types have a relatively low T_c (~25°K max) and rather conventional physical properties, in contrast to the abnormal properties seen in YBCO and Bi2212. With only one CuO₂ layer per unit cell, they are also much simpler. We have recently succeeded in revealing the Fermi surfaces of NCCO with x=0.15 and and x=0.2. The observed Fermi surface of NCCO is a rounded corner square with strong nesting along the (100) direction and it can not be reproduced theoretically with only nearest neighbor interactions. This suggests that the next nearest neighbor interactions must be considered. We have also observed a change of the Fermi surface that is consistent with a band filling scenario.

This is the first time a Fermi surface change with doping has been observed in a HTSC and this is important to understand the doping behavior of cuprate superconductors.

Packet-based transmission of digitally coded images and video over channels exhibiting loss presents a reconstruction problem at the decoder. Compensation for information lost in transmission is required, via retransmission, redundant transmission, or reconstruction techniques using only the received information. We have developed a frequency-based reconstruction technique for lost transform coefficients of known locations, applicable to both luminance and chrominance components, requiring no redundant transmission of information, and using all received information. This technique performs well under both partial and complete loss of coefficients in a block and can also satisfactorily reconstruct multiple adjacent affected blocks. The video reconstruction technique was evaluated using three types of real-time video sequences. Compared with other reconstruction methods, our technique reconstructs both spatial information and low movement sequences at very low hardware cost, and the reconstructed image quality is excellent, with minimal errors occurring in fast movement and detailed sequences.

Unit: 1

**TITLE: Investigation of Transport
In Quantum Dots**

Principal Investigator: J. S. Harris, Jr.

Graduate Students: C. Duruöz and M. Youssefmir

1. Scientific Objectives

With recent developments in e-beam lithography techniques, it has become possible to fabricate devices with minimum feature sizes on the order of 100nm. Two dimensional electron gas (2-DEG) structure implemented with GaAs technology is an exceptionally convenient system for studying these nanoscale devices. Because of the future promise that these devices show, we shifted our focus towards them after finishing our Investigation of High Speed Devices.

2. Summary of Research

A particularly interesting nanoscale device is called "Quantum Dot", a submicron region of a semiconductor which confines some number of electrons. In such small quantum dots, one may see significant electron-electron interaction effects, a particular one being called Coulomb Blockade. In the famous experiment done at MIT [Meirav], the tunable density electrons in a single dot were coupled to two semi-infinite leads by tunnel barriers. Then the linear response conductance of the sample was measured as a function of the electron density in the dot. The amazing result was that, the conductance was periodic, the period being proportional to the charging energy required to add a single electron to the dot. This occurs because the tunneling of an electron is "Blocked" due to the "Coulomb" repulsion with the charge already existing in the dot depending on its amount.

Later, the detailed study of the peaks showed that, the amplitude and width of the peaks depend on the quantum levels of the dot [Meir]. Another experiment allowed the study of these levels, in the presence of a magnetic field [McEuen]. Today, the single dots are quite well understood.

Although the temperature to observe Coulomb Blockade is low presently, the scaling of the device by 10 would increase the charging energy by 100 since it depends on the area of the device. This would make it possible to go to temperatures higher than liquid Nitrogen in the very near future, therefore possibility of device applications. Moreover, it is very important to study this fundamental limit of correlated transport of electrons (i.e., one at a time).

2.1 Quantum Dot Arrays

Another system which shows Coulomb Blockade oscillations is the series combination of two Al-Al₂O₃ tunnel junctions. For this material system, one and two dimensional arrays have been fabricated, and very interesting physical effects such as charge soliton transport [Amman] and KTB phase transition [Mooij] were observed. Since the metal islands have very high carrier concentration, the states at the fermi level are very close to, each other, therefore the quantum levels do not have any significant effect in the experiments in contrast to semiconductor quantum dots.

By being motivated with the experiments in Al tunnel junctions, and single dots, we decided to investigate the transport in one and two dimensional arrays of dots in GaAs material system, which is likely to be more prolific because of the possibility of adjusting the electron concentration and quantum levels.

2.2 Physics of Quantum Dot Arrays

2.2.1 One Dimensional Arrays and Charge Solitons

One dimensional arrays show interesting physical effects. In particular, in the intermediate coupling limit which occurs when C_0 is not much smaller than C where C is the capacitance between two dots and C_0 is the capacitance to ground, an excess charge in a dot affects nearby nodes causing some nonuniform charge distribution. This distribution has a width related to the ratio of the above capacitances and behaves like a particle, hence usually called a "Charge soliton" [Amman]. The importance of solitons is due to the fact that they are quite insensitive to the background potential fluctuations, regarding transport (Fig. 1).

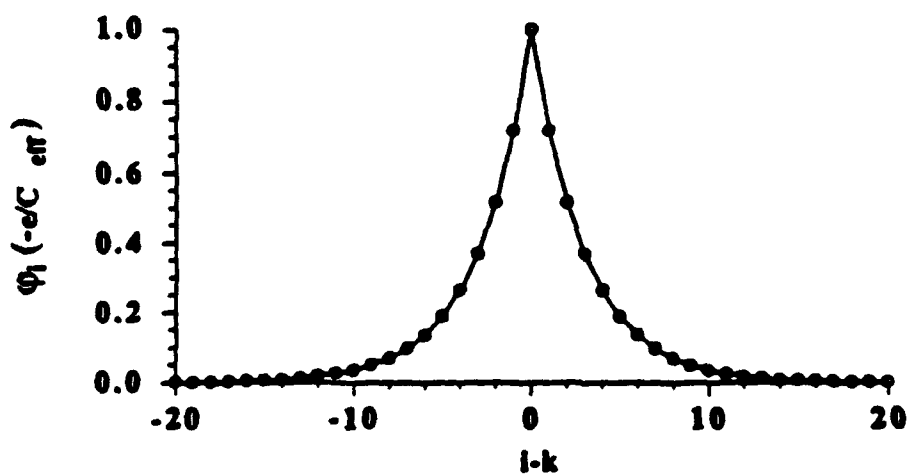
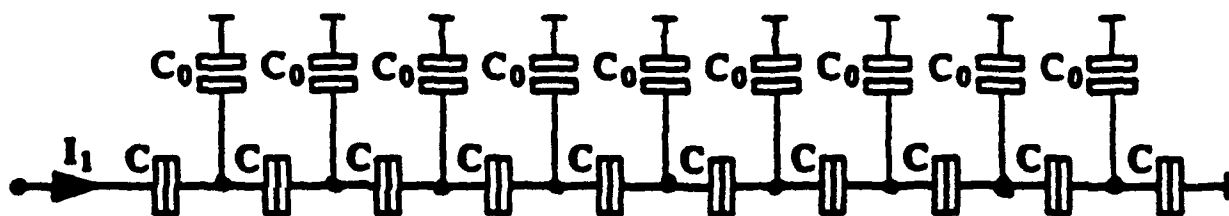


Figure 1 The potential distribution created by a single electron placed on an island k inside an infinite 1D array, i.e., a charge soliton. The "soliton length" M is chosen to be 3. The rings represent the potentials of the dots, the solid line is just a guide for the eye.

2.1.2 Two Dimensional Arrays

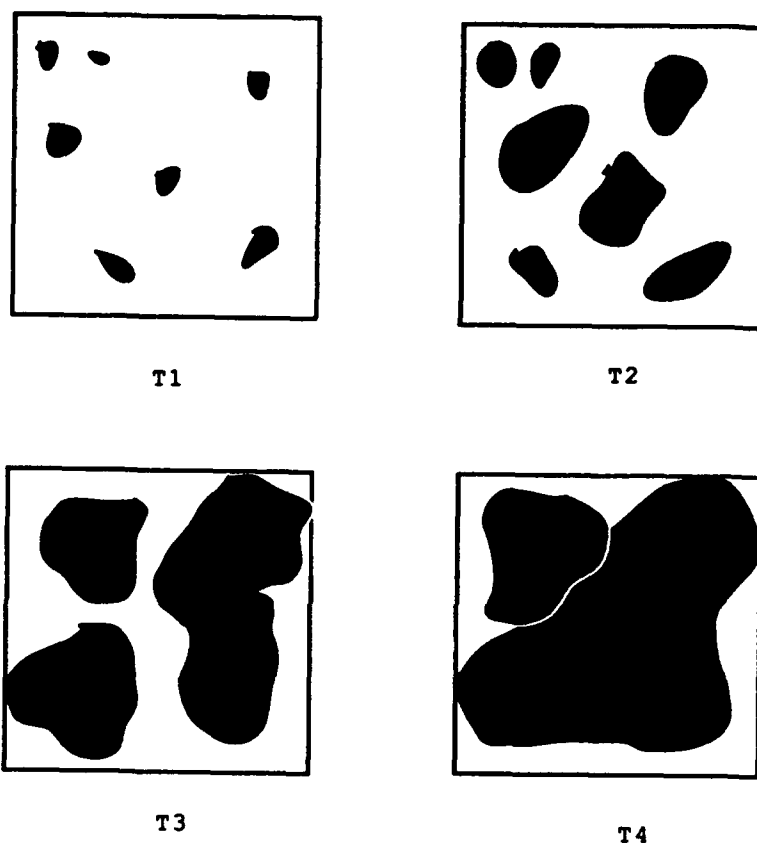
If the array is very regular, and the interactions of the electrons located at second nearest neighboring dots are neglected, the electrons interact with a logarithmic potential [Mooij], therefore it is likely to show KTB (Kosterlitz-Thouless-Berezinskii) phase transition at a critical temperature. One can think of this transition as the unbinding transition of soliton-antisoliton pairs, which are in a bound state for temperatures lower than the critical one. But since the off diagonal interactions are not negligible in semiconductors, the nature of this phase transition is expected to be different or even smeared out.

In the irregular array limit, which is a more realistic consideration, one might expect to see percolating behavior: the number of electrons in each dot will be randomly distributed around some value due to imperfections of e-beam lithography and randomly distributed ionized donor impurities. Moreover the dots will have different energy levels than each other. For a single quantum dot the width and height of a conductance peak depend on the energy levels as well as the temperature (Fig. 2). As the temperature is increased, the conductance between individual dots will increase. There will be different threshold temperatures for each pair of dots. So some region of the sample will be conducting at a lower temperature than the other regions. If there is not a connecting path though, the whole array is still insulating. At a critical temperature, the growing clusters will connect and the sample will be conducting. Computer simulations with temperature dependent 2-D resistor arrays demonstrate the same behavior.

2.3 Implementation of Quantum Dot Arrays

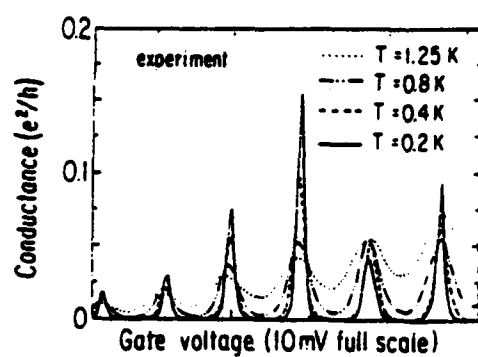
2.3.1 Two Dimensional Electron Gas (2-DEG)

To implement the arrays, high quality two dimensional electron gas (2-DEG) structures were grown by Molecular Beam Epitaxy (MBE), at Stanford Solid State Laboratories. Two different designs were chosen to study different limits: the first design aimed to get highest possible electron mobility with the MBE machine at Stanford. The second design aimed to get low carrier concentrations such as $9 \times 10^{10} \text{cm}^{-2}$. The first design had a 300Å spacer layer and 170Å doped $\text{Al}_{0.34}\text{Ga}_{0.66}\text{As}$ layer. The second one had a spacer of more than 900Å. To decrease the effect of impurities in the samples, several tricks were used: a superlattice buffer which traps the unwanted atoms and makes the 2-DEG grown after few hours cleaner. Outgassing the sources before the growth, that is to heat to sources



$T1 < T2 < T3 < T4$

■ : Conducting regions



Experimental conductance of a narrow GaAs channel with two lithographically defined barriers plotted against gate voltage for four temperatures

Figure 2 Temperature dependence of conductance peaks and percolation.

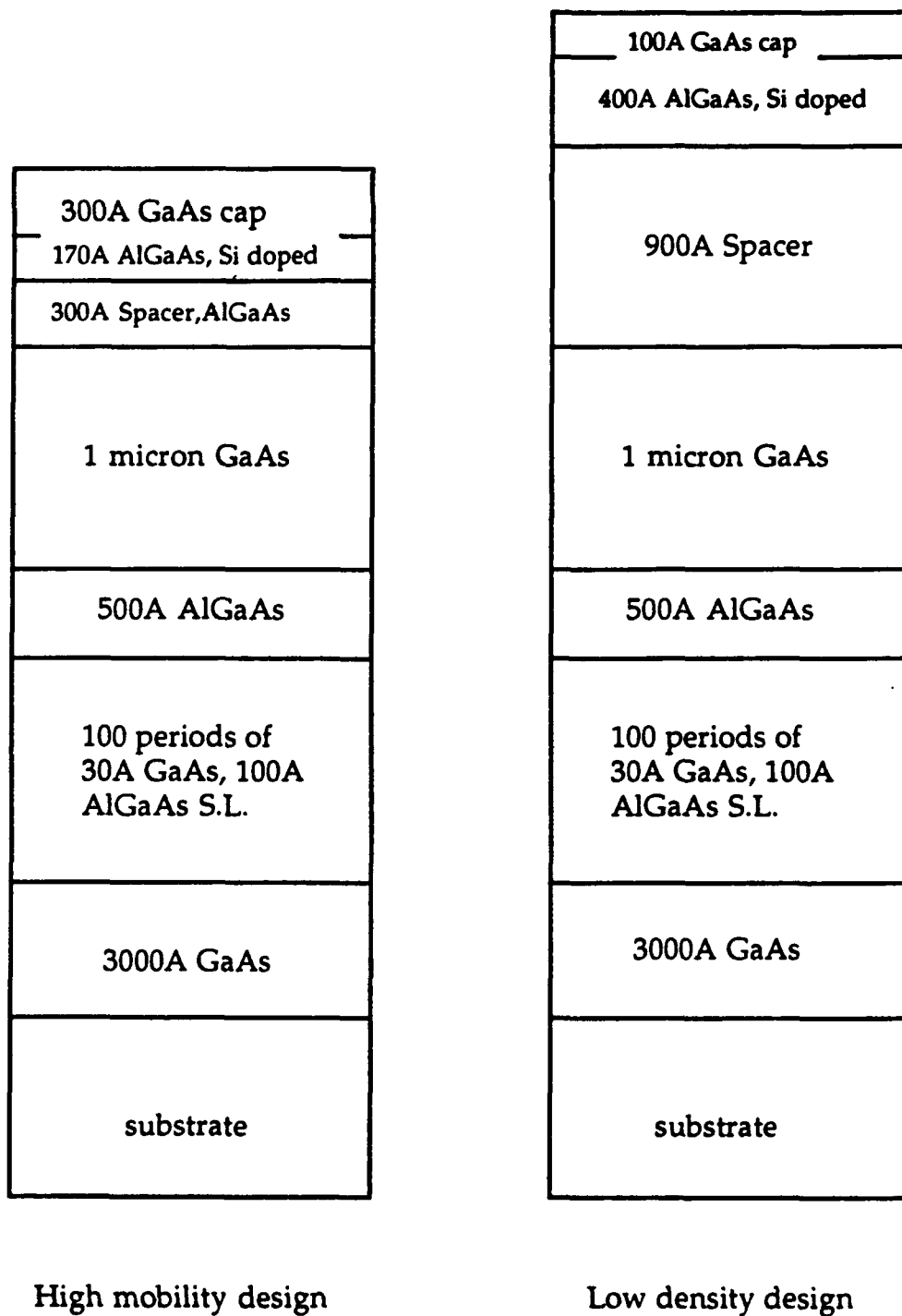


Figure 3 2-DEG structure.

to a higher temperature than the growth temperature to get rid of the impurities in the source. Opening the Al shutter just before the growth, so that it reacts with foreign atoms, and cleans the machine. After around ten runs, we could get several first design wafers with mobilities close to $1,000,000\text{cm}^2/\text{Vs}$, and several low concentration wafers with mobilities around $400,000\text{cm}^2/\text{Vs}$ which is quite a good result for a $9 \times 10^{10}\text{cm}^{-2}$ density 2-DEG (Fig. 3). These mobility results were confirmed by several Quantum Hall Effect measurements, in Ginzton Laboratories at Stanford.

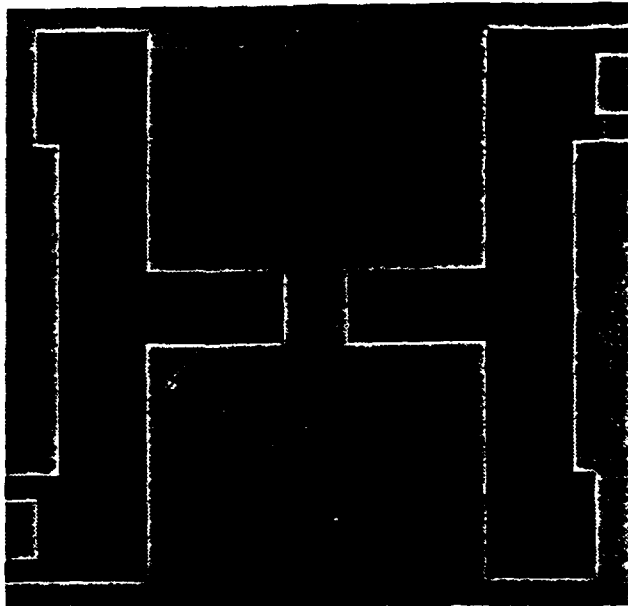
2.2.2 Processing and Electron-beam Lithography

The next step has been the fabrication of the arrays. For making the devices, first we used a standard optical lithography procedure to define the mesas for devices, and put the ohmic contacts which consisted of 250Å Ni, 150Å AuGe alloy and 2200Å Au, annealed at 440°C for 2 mins. This makes the devices ready for the e-beam lithography procedure. Another interconnect layer was evaporated also, to define the coarse gates, and to have e-beam alignment marks ($4 \times 4\text{mm}^2$ squares at some specific locations) on the wafer which are essential for alignment of the e-beam pattern to a partially processed sample.

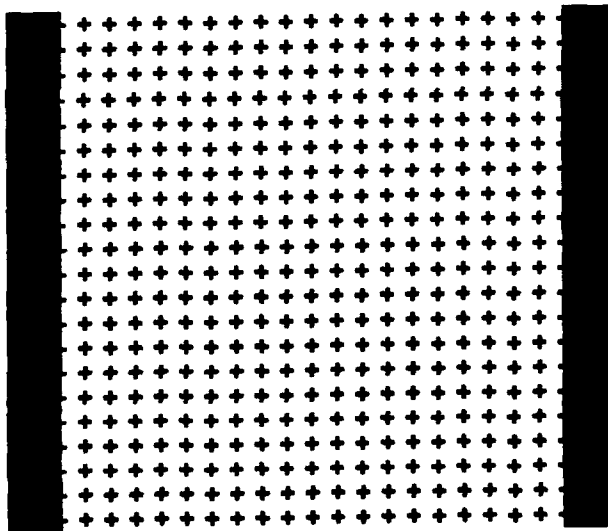
For e-beam lithography, first we used a standard pattern making software called DW-2000, to make a GDSII file which is the input file format for the e-beam machine which will draw the patterns. The patterns consist of plus sign like features repeated with a spacial frequency of 0.8mm (Fig. 4). These regions will be exposed by the e-beam and then will be etched to confine the electrons in the space between them. Since the coupling between the dots depends exponentially on the tunnel barrier thickness, we put arrays with different dimensions, in particular different openings (0.3-0.5mm) to be sure that we get few of the devices with the correct amount of tunneling resistance. In the GDSII file there are fifteen 100×100 , four 20×20 , four 5×100 , five 20×2 , three 100×1 , two 20×10 arrays. Moreover there are two e-beam patterned hall bars as test structures. Each array has two control devices very close to it, one double barrier, one single barrier. They will be controlled by the same gate, and therefore we will be able to know the transparency of each barrier in the array by looking at the nearby single barrier.

3. Future Work

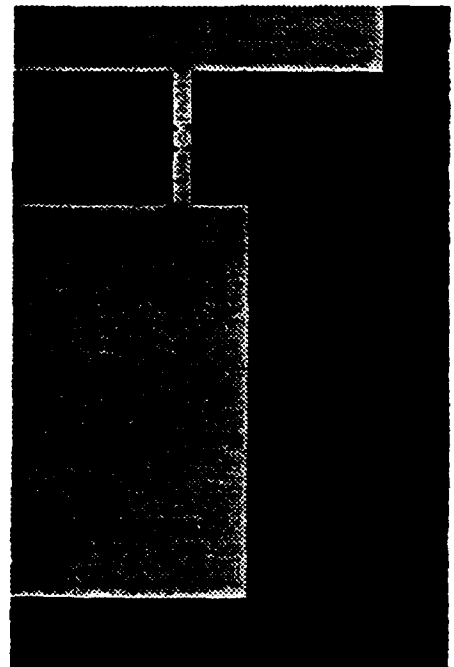
The project is at the electron beam lithography step. Three pieces from different wafers were optically processed, and are ready for



View of a device, grey:mesa, black:e-beam exposed regions (etched)
(array not seen at this magnification since the features are small)



20x20 array



Control device: double barrier

Figure 4 Electron beam patterns.

electron beam lithography. We are expecting to get the exposure done within the next two months. Meanwhile, we are collaborating with the U.C. Berkeley Physics department regarding the usage of their dilution refrigerator for the measurements. A standard AC lock-in technique will be used for the measurements.

Meanwhile we are planning to get more 2-DEG wafers to make more devices, and later study the transition regime between 1-D and 2-D.

4. References

- [Meirav] U. Meirav, M. A. Kastner, S. J. Wind, *Phys. Rev. Lett.* **65**, 771 (1990).
- [McEuen] P. L. McEuen, et al, *Phys. Rev. Lett.* **66**, 1926 (1991).
- [Meir] Y. Meir, N. S. Wingreen and P. A. Lee, *Phys. Rev. Lett.* **66**, 3048 (1991).
- [Mooij] J. E. Mooij, et al, *Phys. Rev. Lett.* **65**, 645 (1990).
- [Amman] M. Amman, et al, *Phys. Lett. A* **142**, 431 (1989).

JSEP Supported Publications

1. W. S. Fu, G. R. Olbright, J. F. Klem and J. S. Harris, Jr., "Optical Gain and Ultrafast N Reponse in GaAs/AlAs Type-II Quantum Wells," *Applied Physics Letters*, 61(14), October 1992.
2. D. Costa and J. S. Harris, Jr., "Low-frequency Noise Properties of N-p-n AlGaAs/GaAs Heterojunction Bipolar Transistors," *IEEE Transactions on Electron Devices*, 39(10), October 1992.

JSEP Supported Ph.D Thesis

1. Winston Su-Kee Fu, "Femtosecond and Quasi-Steady State Optical Non-Linear Physics of GaAs/AlAs Type II Quantum Wells," August 1992.

Unit: 2

**TITLE: Physics and Applications of Ultra-small
Materials Structures**

SENIOR INVESTIGATOR: R. F. W. Pease

RESEARCH STUDENT: H. Liu

Introduction

The recent observations of room temperature visible photoluminescence [Canham] and electroluminescence [Koshida] from electrochemically etched porous silicon has rekindled interest in the potential of using silicon as an optical material. In addition, it raises many fundamental questions regarding the possibility of the luminescence originating from the sub-5 nm silicon crystallites in porous silicon. Theoretical studies of idealized Si quantum wires based on tight-binding calculations [Sanders] and first principles local density calculations [Hybertsen] also indicate that the confinement induced transition from an indirect to a direct material is a valid concept. However, to begin to understand this process quantitatively, parameters relating to the intrinsic properties of Si quantum wires are required. These properties can best be obtained from well controlled and ordered nanostructures, which are the subject of this study.

Scientific Objective

The main scientific objective of this study is to obtain precise structural, electrical, and optical information from well behaved Si nanostructures so that one can begin to understand their intrinsic electronic structures.

Potential Applications

- The ultimate technological goal is to explore the potential of Si optoelectronic devices.
- The processing and characterization technologies developed in this study will be useful in the future implementation of ultra large scale integration (ULSI).

- The nanostructures developed in this study can be used to facilitate the study of the Si oxidation mechanisms [Liu(1)] and the nucleation mechanisms of diamond growth on Si [Dennig].

Work Accomplished

- We have successfully fabricated high aspect ratio Si nano-wires using a combination of high resolution electron beam lithography and anisotropic reactive ion etching. It was also found possible to engineer the profile of the Si nano-wires by intentional introduction of Chromium (Cr) into the Chlorine (Cl₂) reactive ion etching chamber.

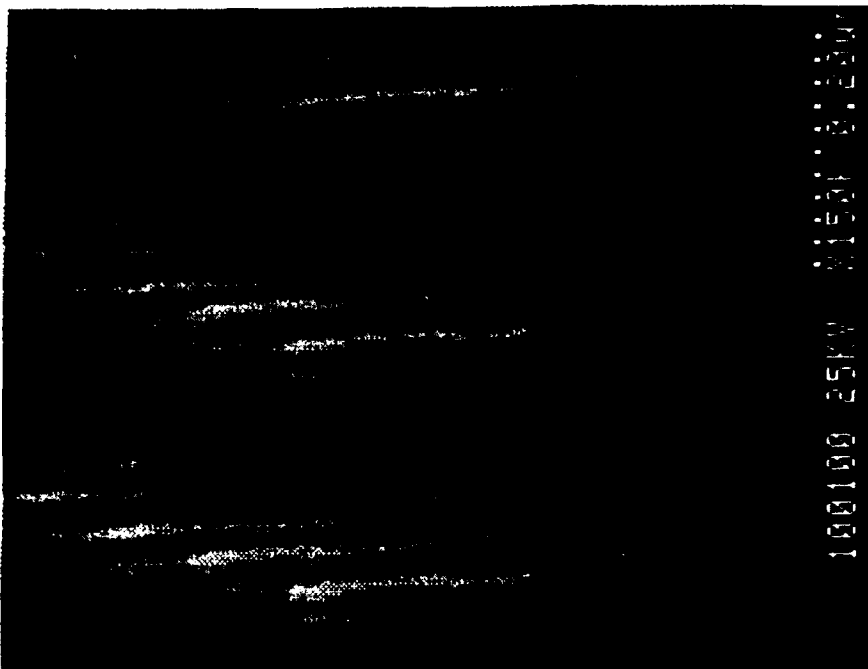


Fig. 1 SEM micrograph showing the high aspect ratio (35 nm in diameter and 400 nm in height) of Si nano-pillars.

- With thermal oxidation, the Si core diameter can be further reduced to below 10 nm. A transmission electron microscopy (TEM) technique was developed to obtain the precise structural information of the Si nano-wires. Lattice images of the Si cores surrounded by oxide skins can be easily obtained with this technique.

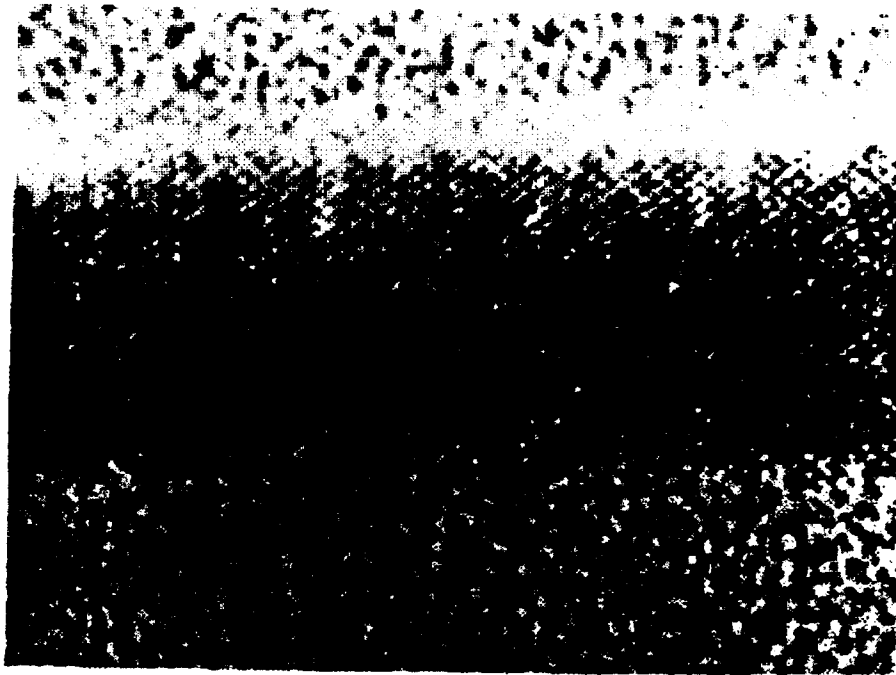


Fig. 2 TEM micrograph showing a portion of a 8 nm crystalline Si core surrounded by amorphous oxide skin.

- **Because of the nondestructive nature of the TEM technique, we were able to tract the evolution of an individual identical Si nano-wire through a series of oxidation steps and to characterize the oxidation process of these nano-wires.**
- **Photoluminescence signal was obtained from these Si nano-wires [Liu(1)].**
- **Device structures based on making contacts to polyimide planarized p-i-n Si nano-wires were fabricated. Preliminary current-voltage characteristics indicate that reliable contacts can be made to these nanostructures [Liu(2)].**

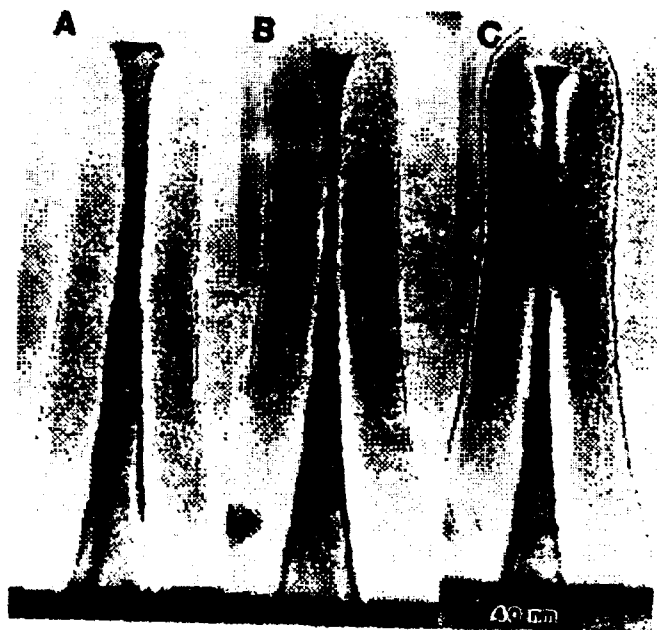


Fig. 3 A series of TEM micrographs showing an identical Si pillar going through a series of thermal oxidation steps at 850 C. (A: 3 hours; B: 4 hours; C: 5 hours)

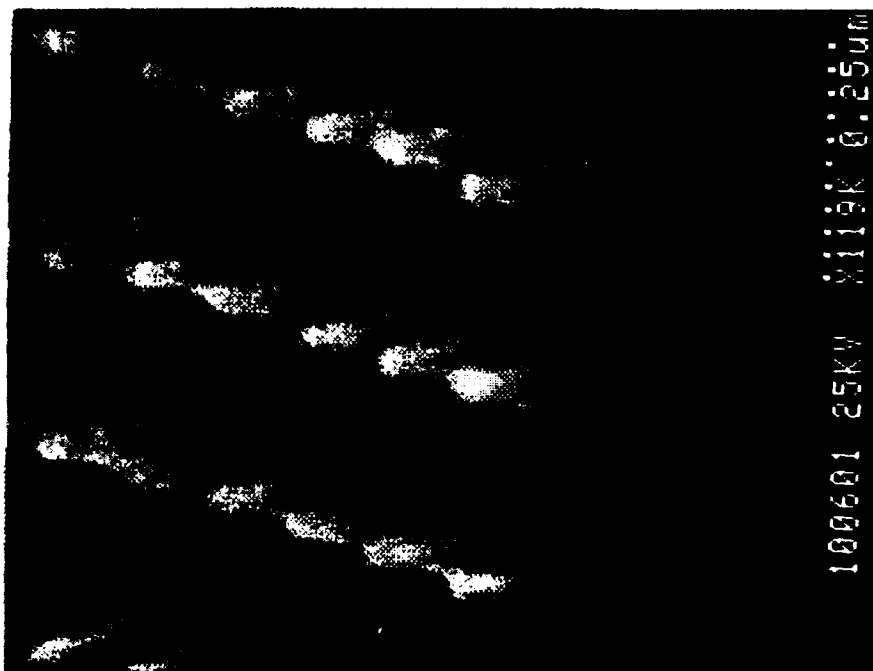


Fig. 4 SEM micrograph showing the tops of Si pillars protruding out of the polyimide planarizing medium.

Work in Progress

- Techniques are being developed to fabricate and to align to a small number of isolated Si nano-wires so that one can obtain the intrinsic electrical properties of these structures without the statistical variations.

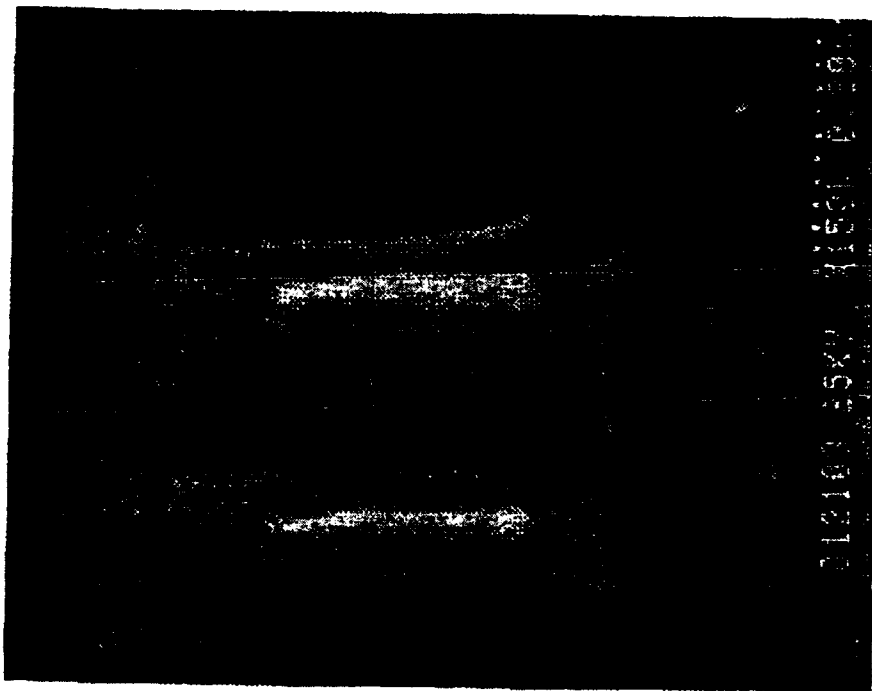


Fig. 5 SEM micrograph showing the ability to construct and locate a small number of isolated Si nano-pillars.

- Once the contacting techniques are perfected, transparent conductor will be used to make the electrical contacts to these nanostructures. The optical properties of the light emitting device structures will be characterized.
- More elaborate TEM sample preparation techniques are being developed to obtain even higher resolution images of the crystalline Si core in the nano-wires. This allows one to detect defects due to the processing and to begin uncover the oxidation mechanisms in these nanostructures.
- The necessity of making fine scratches on clean Si prior to chemical vapor deposition (CVD) growth of diamond film on Si has raised the issue regarding the effect of surface topography on the

nucleations of diamond growth [Dennig]. The ability of fabricating Si pillars of various sizes allows one to create controlled surface topography and the diamond nucleations in a systematic fashion. Moreover, the ability to grow high quality diamond on the Si nano-wires also has device implication. Since diamond is a high band gap material, it may be an efficient emitter for injecting carriers into the high energy quantized states in the nano-wires. Efforts are now being made to perform CVD diamond growth on the fabricated nanostructures.

References

- [Canham] L. T. Canham, *Appl. Phys. Lett.* **57**, 1046 (1990).
- [Dennings] P. A. Dennig, H. Shiomi, D. A. Stevenson, N. M. Johnson, *Thin Solid Films*, **212**, 63-67 (1992).
- [Hybertsen] M. S. Hybertsen and M. Needels, *MRS 1992 Fall Proceedings Symposium F* (1992).
- [Koshida] N. Koshida and H. Koyama, *Appl. Phys. Lett.* **60**, 347 (1992).
- [Liu(1)] H. I. Liu, N. I. Maluf, R. F. W. Pease, D. K. Biegelsen, N. M. Johnson, F.A. Ponce, *J. Vac. Sci. Technol. B*, **10** (6), 2846-2850 (1992).
- [Liu(2)] H. I. Liu, D. K. Biegelsen, N. M. Johnson, F. A. Ponce, N. I. Maluf, R. F. W. Pease, *MRS 1992 Fall Proceedings Symposium F* (1992).
- [Sanders] G. D. Sanders, C. J. Stanton, Y. C. Chang, *MRS 1992 Fall Proceedings Symposium F* (1992).

JSEP Supported Publications

1. H. I. Liu, N. I. Maluf, R. F. W. Pease, D. K. Biegelsen, N. M. Johnson, F. A. Ponce, "Oxidation of Sub-50 nm Si Columns for Light Emission Study", *J. Vac. Sci. Technol. B*, **10** (6), 2846-2850 (1992).
2. H. I. Liu, D. K. Biegelsen, N. M. Johnson, F. A. Ponce, N. I. Maluf, R. F. W. Pease, "Silicon Quantum Wires: Oxidation and Transport Studies", *MRS 1992 Fall Proceedings Symposium F* (1992).

Unit: 3

TITLE: Reactive Ion Profiling of Heterostructures

PRINCIPAL INVESTIGATOR: C. R. Helms

GRADUATE STUDENTS: T. Beerling and R. Keller

Scientific Objectives

The initial objective of this work was to determine the surface chemistry associated with the interaction of reactive ions with GaAs and other semiconductor surfaces. The objective of this work is to determine how oxygen ion irradiation effects the quality of diamond surfaces. In-situ Electron Energy Loss Spectroscopy (EELS) and Auger Electron Spectroscopy (AES) are being employed in this investigation. In addition, we are continuing work begun on the previous program on metal-GaAs interfaces [Kniffin, '90, Kniffin, '91, Weiss, '90]. Here we are developing an oxidation process for the selective removal of Ga from the near surface region of GaAs and use this process to study stoichiometric effects on dopant diffusion and activation.

Key results include determination of the mechanisms of oxygen ion etching of diamond surfaces, observation of an "ideal" oxygen terminated diamond surface and the use of GaAs selective oxidation for dopant incorporation and activation.

The Effects of Oxygen Ion Bombardment to Damaged and Undamaged Diamond Surfaces

We have found that 200 eV oxygen ions will remove damaged (non-diamond) layers of diamond, leaving a sp^2 free surface. The oxygen bombardment provides surface atomic oxygen with the dissociation of oxygen ion molecules, or by direct impingement of atomic oxygen ions, which etch carbon. Any damage created by the 200 eV oxygen ion bombardment is quite small in relation to the etching rate, and also contained to a very shallow region, due to the small ion range. Therefore, "good" diamond surfaces are obtained. Higher energy oxygen ions (500 eV or higher) do not leave such ideal surfaces. Etching still occurs for these higher energy irradiations. However, these physical sputtering/knock-on processes also creates a higher rate of damage, apparently faster than the oxygen ions can chemically etch or physically sputter.

Figure 1 shows the EELS spectra for (100) natural diamond, taken in the reflection mode at a primary energy of 500 eV. The spectra is indicative of "good" diamond, with bulk and surface plasmons at 35 eV and 22.5 eV, respectively. Peaks are also seen at 9 eV and 14.5 eV, attributed to electronic transitions [Lurie]. EELS spectra for synthetic diamond was seen to be essentially identical to the spectra taken for (100) natural diamond. Figure 1 also presents EELS spectra for a surface with 1keV Xe⁺ ions. damaged For comparison, the EELS spectra of highly ordered pyrolytic graphite (HOPG) is included. The bulk diamond plasmon is barely distinguishable.

The EELS spectra for a damaged synthetic diamond exposed to 200 eV oxygen ions is shown in Fig. 2. During oxygen ion exposure, the π -plasmon shifts from 4.5 eV to 5.5 eV. This can be explained as oxygen ions etching the most damaged regions, which should also be the least dense due to displacement of carbon atoms. With continued oxygen ion exposure, the π -plasmon is seen to completely disappear, as all the damaged diamond has been removed. The final spectra (1.2×10^{17} cm⁻² dose of oxygen ions) shown is similar to a "good" diamond surface.

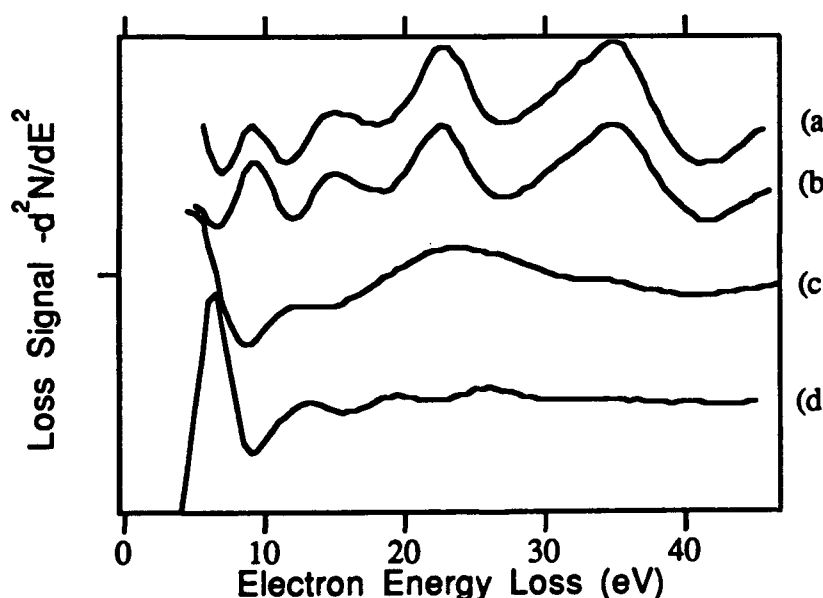


Figure 1 EELS spectra of a): Natural (100) diamond, b): H₂SO₄ cleaned synthetic diamond, c): Xe damaged synthetic diamond, 1 keV, 5×10^{15} cm⁻² fluence, d): HOPG.

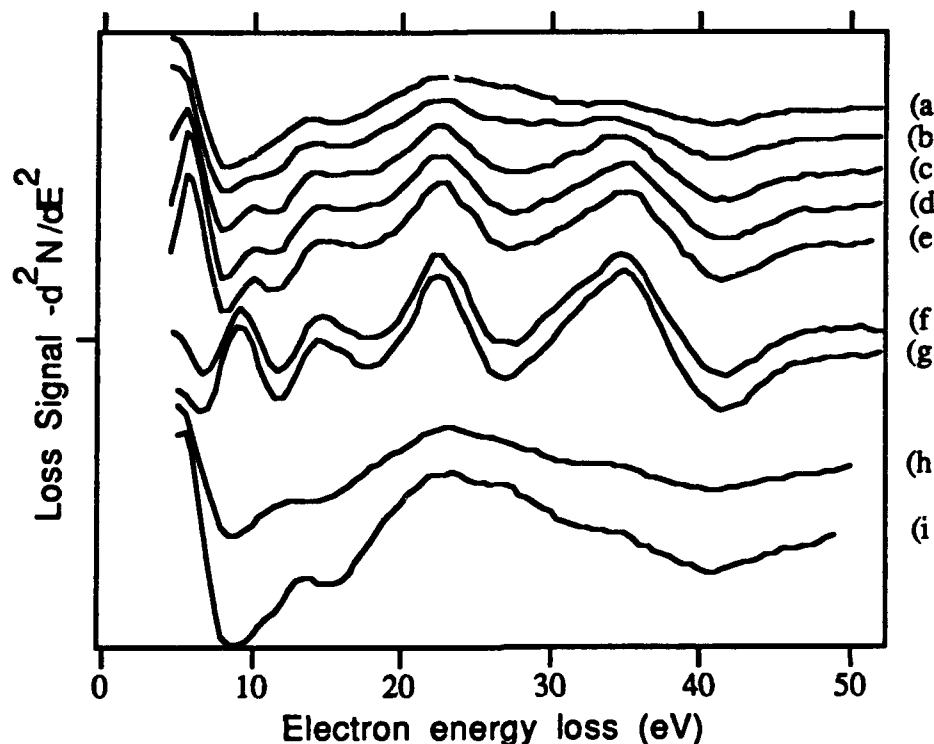


Figure 2 EELS spectra of initially Xe damaged (with $5 \times 10^{15} \text{ cm}^{-2}$ Xe^+ ions) synthetic diamond exposed to 200 eV and 1keV oxygen ions. For the 200 eV oxygen ion irradiation: a): Xe damaged, b): $4.8 \times 10^{15} \text{ cm}^{-2}$ exposure, c): $1 \times 10^{16} \text{ cm}^{-2}$, d): $2 \times 10^{16} \text{ cm}^{-2}$, e): $4 \times 10^{16} \text{ cm}^{-2}$, f): $8 \times 10^{16} \text{ cm}^{-2}$, g): $1.2 \times 10^{17} \text{ cm}^{-2}$. For the 1keV oxygen ion irradiation: h): Xe damaged, i): $1.1 \times 10^{17} \text{ cm}^{-2}$ exposure.

As oxygen ion energy increases, the degree of recovery of "good" diamond spectra decreases. At some point, no improvement is seen in the EELS spectra after oxygen irradiation. This can be seen in Fig. 3; for 1000 eV oxygen ion exposure, essentially no improvement is seen whatsoever after an appreciable fluence of oxygen ions.

We are also continuing studies on 2b natural diamond (naturally conducting), where we find qualitative agreement with the results obtained for synthetic diamond. Work is being continued on synthetic diamond primarily to determine how ion irradiation of diamond surfaces with oxygen and inert ions effect I-V characteristics of metal diamond interfaces, as it has been shown that inert ion irradiation will change I-V characteristics [Tachibana].

To determine the surface structure of oxygen ion irradiated diamond surfaces, Low Energy Electron Diffraction (LEED) will be performed on (100) natural 2b diamond surfaces. Surface reconstructions, crystalline integrity as a function of incident oxygen ion energy, as well as recovery of crystallinity from damaged surfaces will be investigated.

Selective Oxidation of Gallium Arsenide

We have developed a selective oxidation process for the removal of Ga from the near surface region of GaAs. We have applied this process to the study of diffusion of Ge and Si from deposited surface layers. Our results show how the removal of gallium from the GaAs substrate can be controlled with the selective oxidation process, thereby controlling the diffusion of Ge from a deposited layer via Ga vacancy formation. Initial results for diffusion of Si from a deposited layer using our selective oxidation process show promise for the creation of n+ doped GaAs with electron concentrations greater than 10^{19} cm^{-3} .

In the selective oxidation process, a material is annealed in a $\text{H}_2\text{O}/\text{H}_2$ ambient, where the partial pressure ratio of water to Hydrogen, $P(\text{H}_2\text{O})/P(\text{H}_2)$, can be varied. The adjustment of the partial pressure ratio allows for the oxidation of elements whose oxides have large free energies of formation while preventing oxidation of other elements. This technique was first successfully used for the selective oxidation of Si in the presence tungsten [Kobyashi]. The selective oxidation of Ga in an GaAs system was previously demonstrated in this lab [Weiss] for the formation of metal-arsenide contacts to GaAs free of gallium containing compounds (i.e. metal-gallides).

We have investigated the use of selective oxidation to control the stoichiometry of the near surface region by removal of Ga from the substrate in the form of gallium oxide. To create an arsenic rich region at the surface, the GaAs was capped with a thin SiO_2 layer (250Å), allowing diffusion of gallium to the surface and H_2O and H_2 to the substrate while preventing liberation of arsenic. These structures were then oxidized using either water vapor in an Ar carrier or a mixture of that and hydrogen. The water vapor pressure in the Ar was regulated between 20 and 250 Torr by controlling the water bubbler temperature.

We have characterized the oxidation process using Auger sputter profiling. Fig. 3 shows three Auger sputter profiles which clearly show the effect of varying the partial pressure ratio on the oxidation of GaAs with a thin SiO₂ encapsulant. In all cases, the 250Å SiO₂ samples were annealed for 30 min. at 700°C. In Fig. 3(a), the partial pressure ratio of $P(H_2O)/P(H_2) = 0.7$. In this case, a large amount of gallium has formed a layer of gallium oxide between the SiO₂ and GaAs. A large excess of arsenic is clearly seen under the gallium oxide. For temperatures $\leq 700^\circ\text{C}$, oxidation of the 250Å SiO₂ samples produced a gallium oxide below the SiO₂, indicating that the indiffusion of water and hydrogen was dominant. For temperatures above 900°C, the gallium oxide forms on top of the SiO₂, indicating that outdiffusion of the gallium was the dominant kinetic process. In Fig. 3(b), $P(H_2O)/P(H_2) = 0.35$, and the amount of oxide formed has been reduced. In Fig. 3(c) $P(H_2O)/P(H_2) = 0.19$. No oxide is found between the SiO₂ and the GaAs substrate.

We have applied the selective oxidation process to study the effect of different stoichiometric conditions on diffusion and activation of column IV dopants. Under conditions which do not produce Ga vacancies in the underlying substrates, diffusion of Si and Ge from a deposited surface layer is very slow [Chang]. However, the diffusivity increases orders of magnitude under conditions which produce an arsenic rich interface, such as exposure to As overpressures [Antell], annealing with Ga permeable encapsulants such as SiO₂ [Greiner], or As doping of the deposited Si [Kavanagh]. However, even though atomic concentrations of nearly 10^{20} cm^{-3} can be achieved these methods, electrically active dopant concentration have not exceeded $\sim 5 \times 10^{18} \text{ cm}^{-3}$ for any of the above methods. This is consistent with results for ion implantation as well as MBE and MOCVD, where electrically active n-type dopant concentrations are always $< 10^{19} \text{ cm}^{-3}$.

The first structure we investigated to look at dopant effects consisted of a thin (35Å) Ge layer capped by 250Å of sputter deposited SiO₂. Figure 4 shows an Auger sputter profile of the SiO₂/Ge/GaAs structure annealed for 30 min. at 650°C under with $P(H_2O)/P(H_2) = 0.75$. A gallium oxide layer has formed beneath the SiO₂, with the excess As lying below, in the same region as the Ge. The As and Ge have probably reacted to form GeAs.

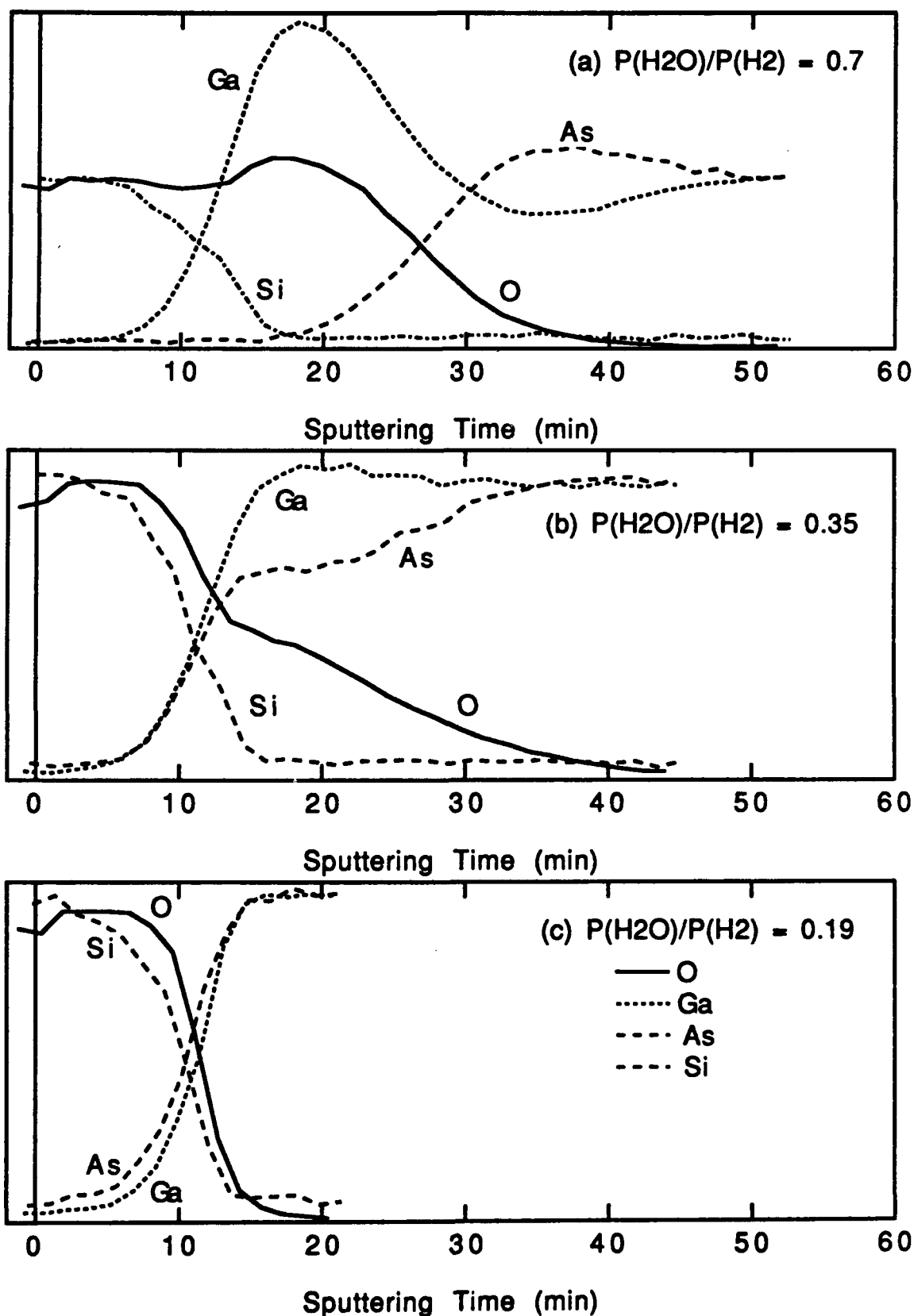


Figure 3 Auger Sputter profiles showing the effect of water/hydrogen partial pressures on the oxidation of GaAs with a 250Å SiO₂ encapsulant at 700°C for 30 min.

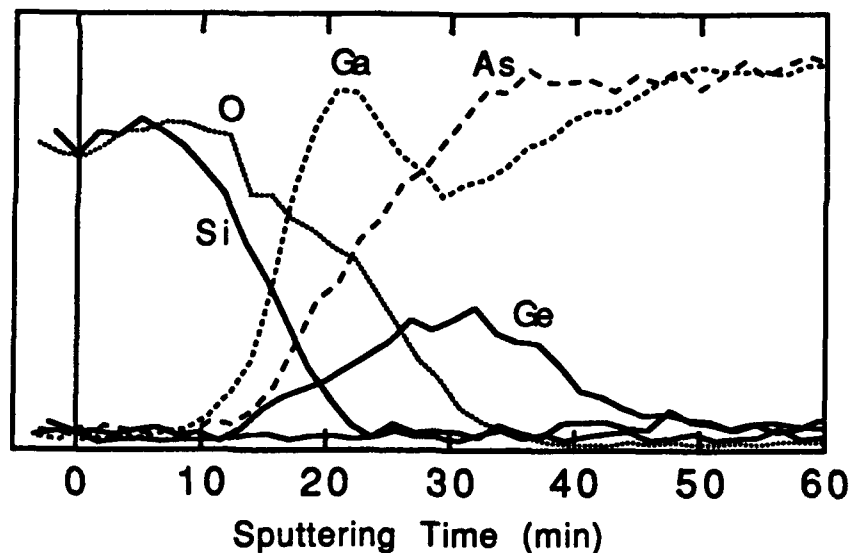


Figure 4 Auger sputter profile of oxidized SiO₂/Ge/GaAs

In order to determine the effect of excess As on diffusion of Ge, samples oxidized at 650°C for 30 minutes under various water/hydrogen partial pressures were annealed at 900°C for 2.5 min. in N₂. The second higher temperature anneal was needed to drive in the Ge dopants, as no diffusion was observed for $T < 750^{\circ}\text{C}$. Figure 5 shows the total electrically active dopants diffused into the substrate as measured by a Van der Pauw technique vs $P(\text{H}_2\text{O})/P(\text{H}_2)$ during the 650°C oxidation step. For $P(\text{H}_2\text{O})/P(\text{H}_2) \leq 0.025$, no diffusion of dopants was detected. For $P(\text{H}_2\text{O})/P(\text{H}_2) > 0.16$, the diffused electron concentration increases until it saturates at $P(\text{H}_2\text{O})/P(\text{H}_2) = 0.4$.

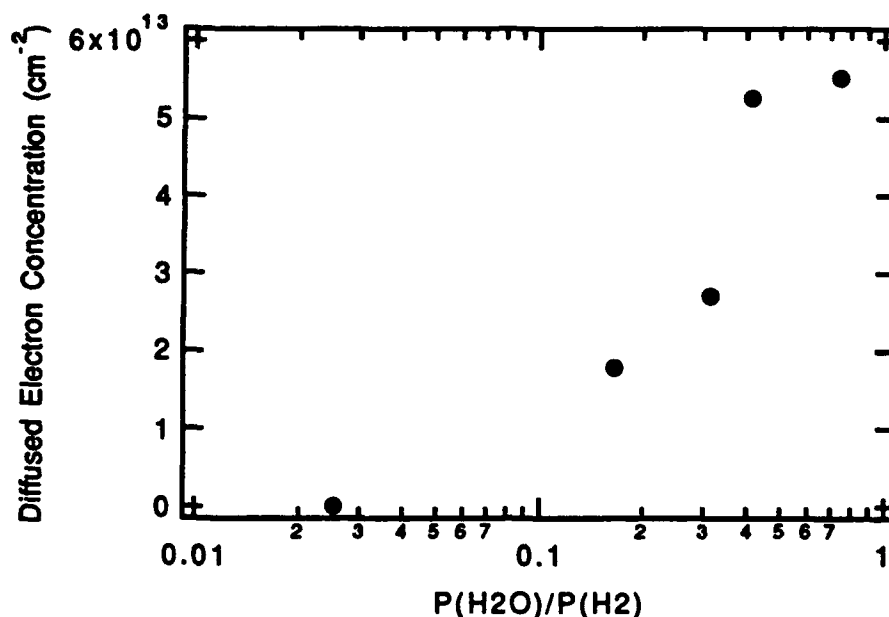


Figure 5 Diffused electron concentration vs partial pressure ratio for SiO₂/Ge/GaAs structure oxidized at 650°C with 900°C anneal.

Further investigations of the diffusion and activation processes from surface layers of Si and Ge will be performed. Specifically, experiments similar to the above work are currently being done for Si layers in order to explain the mechanisms involved in the diffusion and high activation of the rapid thermal oxidation process used by [Sadana]. Initial results indicate doping concentrations in excess of 10^{19} cm^{-3} can be obtained under certain ambient conditions, higher than that previously obtained by Sadana for oxidation without a subsequent rapid thermal anneal. Also, the effect of stoichiometry on the activation of ion implanted GaAs will be studied using the selective oxidation process.

References

- [Antell] G. R. Antell, *Solid State Electronics* **8**, 943 (1965).
- [Chang] J. C. P. Chang and K. L. Kavanagh, *Mat. Res. Soc. Symp. Proc.* Vol. **224**, 1991.
- [Greiner] M. E. Greiner and J. F. Gibbons, *Appl. Phys. Lett.* **47**, 1208 (1985).
- [Kavanagh] K. L. Kavanagh and C. W. Magee, *Can. J. Phys.* **65**, 987 (1987).
- [Kniffin] M. L. Kniffin, C. R. Helms, *J. Appl. Phys.* **68**, 1367 (1990).

- [Kniffin] M. L. Kniffin, Ph.D. Dissertation, Stanford University (1991).
- [Kobyashi] N. Kobyashi, S. Iwata, and N. Yamamoto, *Proceedings of the IEEE International Electron Device Meeting*, Dec. 1984, p. 122.
- [Lurie] P. G. Lurie, J. M. Wilson, *Surf. Sci.*, **65**, 476 (1977).
- [Sadana] D. K. Sadana, J. P. Souza, and F. Cardone *Appl. Phys. Lett.* **57**, 1681 (1990).
- [Tachibana] T. Tachibana, B.E. Williams, J. T. Glass, *Phys. Rev. B*, **45**, 11968, (1992).
- [Weiss] E. Weiss, R. C. Keller, M. L. Kniffin, C. R. Helms, *Appl. Phys. Lett.* **56**, 2557 (1990).
- [Weiss] E. Weiss, R. C. Keller, and C. R. Helms, "Formation of Single-phase MASx Films on GaAs by Selective Wet-Hydrogen Oxidation and Etching" *J. Appl. Phys.* **69** (4), 2623 (1991).

JSEP Supported Publications

1. R. C. Keller and C. R. Helms "Calculations of the Barrier Height and Charge Distribution of a Metal-Dielectric Interface" *J. Vac. Sci. & Technol. A* **10** (4), 775 (1992)
2. M. L. Kniffin and C. R. Helms "Observation of the Instability of Ti_xGa_y alloys with respect to GaAs at Elevated Temperatures" *J. Vac. Sci. & Technol B* **10** (6), 2351 (1992).

Unit: 4

**TITLE: The Electronic Structure and Interfacial
Properties of High Temperature
Superconductors**

PRINCIPAL INVESTIGATOR: W. E. Spicer

JUNIOR FACULTY: Z.-X. Shen

GRADUATE STUDENTS: D. S. Dessau

Scientific Objective:

The objective of this project is two fold: (1) to gain a sound scientific understanding of the bulk electronic structure of the recently discovered high temperature superconductors (HTSCs), and (2) to apply this knowledge to the study of interfaces of the HTSCs with other technologically important materials.

Summary of Research:

(1) Normal state Fermi surface of HTSCs

The normal state of cuprate superconductors exhibits many unusual physical properties, in contrast to the conventional wisdom of the Fermi liquid. This together with (a) the high T_c itself, (b) the proximity to the antiferromagnetic Mott insulator, raise the question whether HTSC in its normal state is a Fermi liquid. Investigation of the experimental Fermi surface is crucial to address this important issue.

a) Bi2212 Fermi Surface. Due to its superior surface quality, Bi2212 is so far the only cuprate compound whose surface structure has been proven to be consistent with the bulk by structural tools such as STM and LEED. We have carried out experiments to measure the Fermi surface over the entire 2D Brillouin zone of this compound [1]. Figure 1 presents selected angle resolved PES spectra from the whole valence band and the near E_F . On the basis of data such as those in Fig. 1, we obtained the experimental Fermi surface as shown in Fig. 2. The various circles represent the k-space locations sampled. The black circles represent the locations where a band is observed to cross E_F , with the solid lines through the black dots representing the experimental

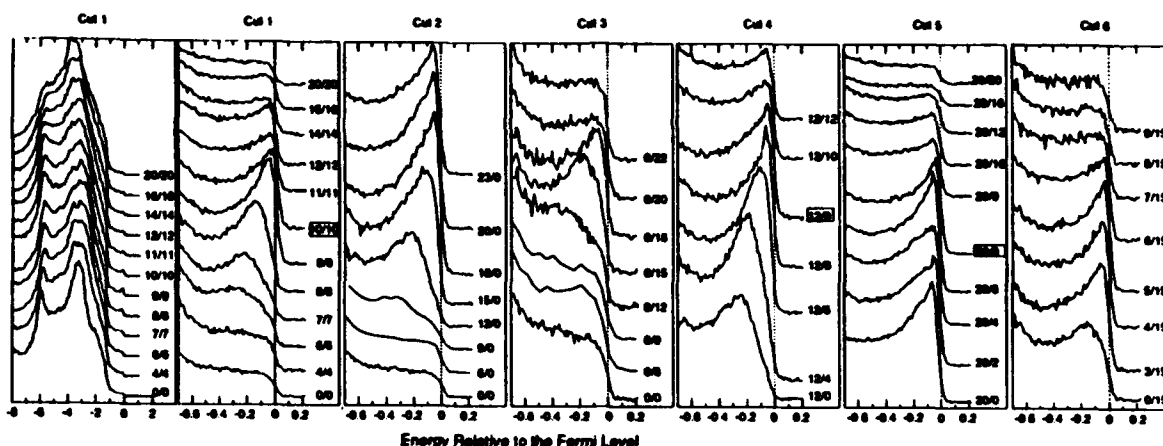


Figure 1 Angle-resolved photoemission data from Bi2212. Cut 1 gives whole valence band data while the other cuts give the high-resolution scans. The numbers marked beside each spectra are angles θ and ϕ , as marked in the grid of Fig. 2

Fermi surface. This experimental Fermi surface can be well described by LDA or tight-binding calculations considering only the two CuO_2 planes per unit cell. The gray circles represent the k -space location (near M) where a flat band very close to E_F has been observed (see for example the top portions of cut 2 and the bottom portions of cut 5). These flat bands arrive from a saddle-point in the band structure which happens to lie very near E_F for Bi2212, and will lead to a strongly anisotropic $N(E_F)$ which will be strongly weighted near M. This may have important implications for understanding many of the anomalous normal state [e.g. Hall effect, susceptibility, thermopower] as well as superconducting state properties [e.g. gap anisotropy (see below), magnitude of the T_c] exhibited by most of the HTSCs. This saddle point is observed at E_F at the corresponding k -space location of YBCO, though is found to be a few hundred meV below E_F in the n-type superconductor NCCO [2].

b) $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_{4+d}$ (NCCO). NCCO and its relatives are particularly interesting because they are the only n-type HTSCs that have been discovered so far. All of the n-types have a relatively low T_c ($\sim 25\text{K}$ max) and rather conventional physical properties, in contrast to the abnormal properties seen in YBCO and Bi2212. With only one CuO_2 layer per unit cell, they are also much simpler, in contrast to the more complicated structures of YBCO and Bi2212, which have two CuO_2 planes as well as the chain or BiO plane layers which may be electrically active. These properties of NCCO reveal the urgent need to study its electronic structure and Fermi surface.

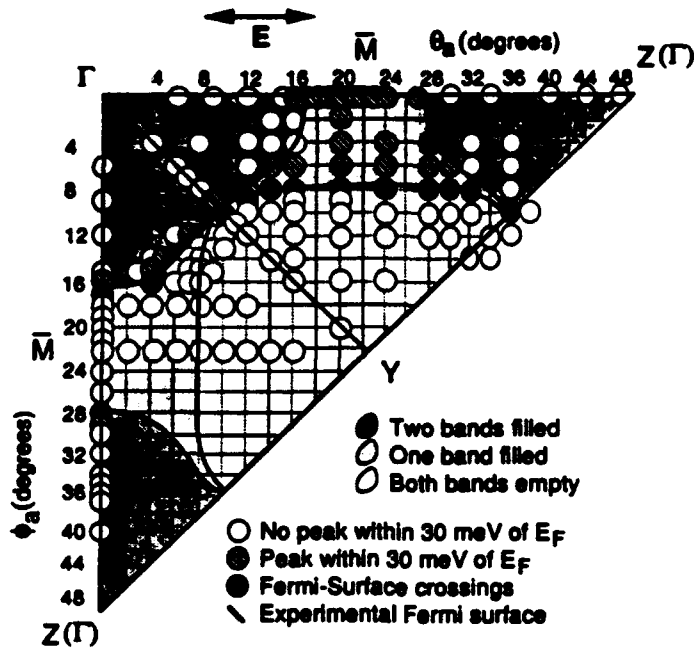


Figure 2 Experimentally measured Fermi surface with the E field of photon being horizontal, as illustrated. More Fermi surface crossings were observed with the E field being vertical. The lines are results from a LDA calculation without considering the mixing of Bi-O state and Cu-O states near E_F .

Technically, NCCO is a very difficult system to study so that earlier studies have failed to reveal structures near E_F . As shown in Figs. 3 and 4, we have recently succeeded in revealing the Fermi surfaces of NCCO with $x=0.15$ and $x=0.2$ [2]. The observed Fermi surface data points (shaded in gray) agree very well with the results of a LDA calculation (solid line). The Fermi surface of NCCO is a rounded corner square with strong nesting along the (100) direction. The observed Fermi surface can not be reproduced theoretically with only the nearest neighbor interactions included. This suggests that the next nearest neighbor interactions should be considered. As shown in Fig. 5, we have also observed a change of the Fermi surface that is consistent with a band filling scenario. This is the first time a Fermi surface change with doping has been observed in HTSC and this is important to understand the doping behavior of cuprate superconductors. Finally, as mentioned in (a) above, the saddle point which gives rise to the flat bands at E_F in Bi2212 and YBCO is observed to be ~ 300 meV below E_F in NCCO. The relatively simple Fermi surface as well as the absence of the saddle point at E_F may be important for understanding why many of the "abnormal" properties of Bi2212 and YBCO are more normal in NCCO.

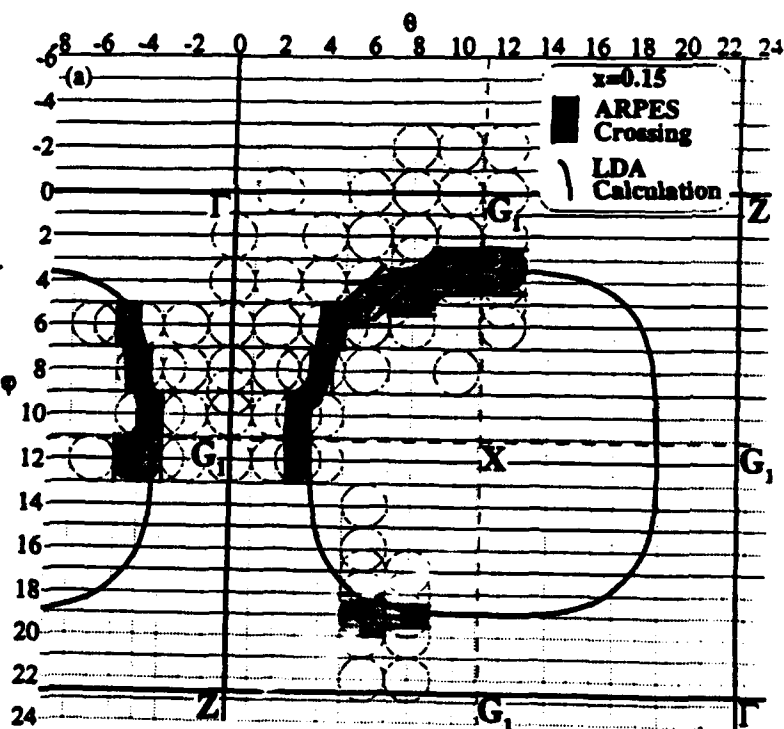


Figure 3 Experimentally measured Fermi surface of NCCO with $x=0.15$ (shaded points) in comparison with results of the LDA calculation (lines).

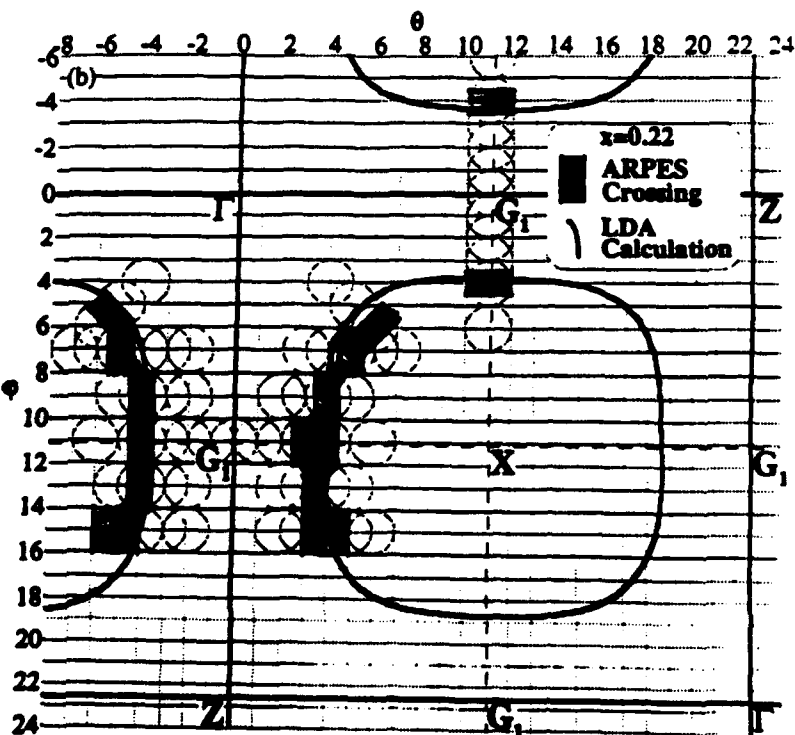


Figure 4 Experimentally measured Fermi surface of NCCO with $x=0.22$ (shaded points) in comparison with results of the LDA calculation (lines).

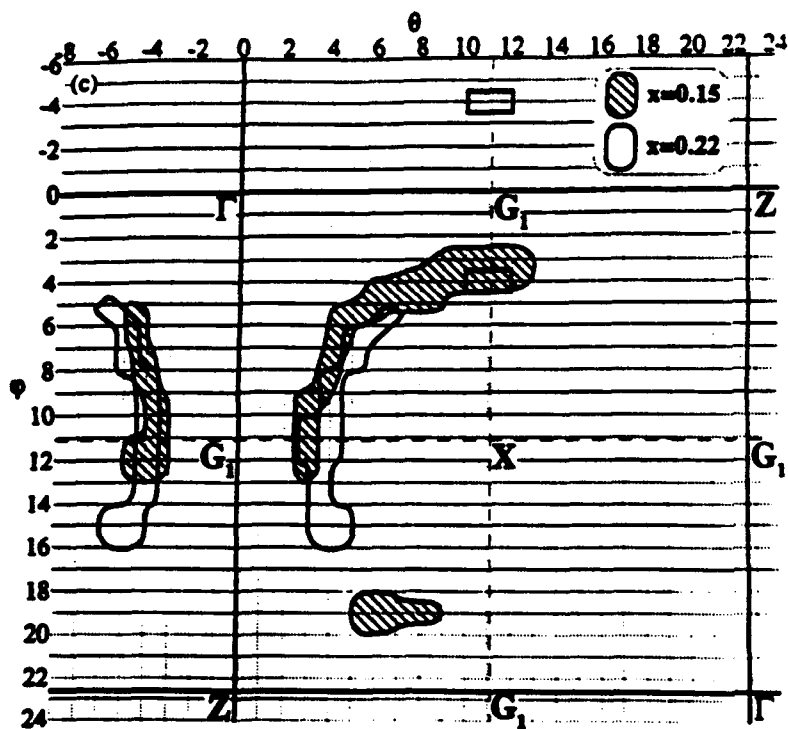


Figure 5 Comparison of experimental Fermi surface of NCCO with $x=0.15$ and 0.22 . As x increases, the Fermi surface surrounding X point shrinks in a fashion consistent with an electron doping scenario.

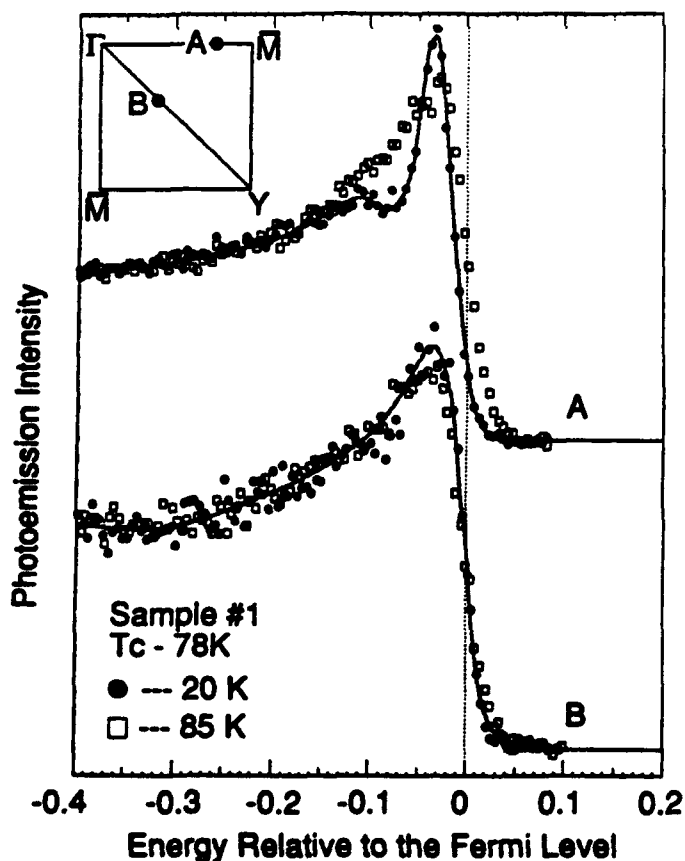


Figure 6 High-resolution photoemission spectra above and below T_c taken at two locations in k -space. The effects of superconducting gap opening is clearly visible in A and is not detectable within our experimental uncertainty at B.

(2) a-b plane superconducting gap anisotropy in Bi2212

Figure 6 presents photoemission spectra above and below T_c of Bi2212 at two different locations in k-space [3]. The spectra taken at k-space location B hardly change above and below T_c , suggesting that the gap is undetectable within our experimental uncertainty (± 2 meV). On the other hand, the spectra taken at A clearly show the effects of the gap opening: the leading edge is shifted to higher binding energy, and a high binding energy dip structure is observed. The shift of the leading edge at A gives an order of magnitude estimate of the magnitude of the gap, and is found to range from approximately 12 to 20 meV depending upon the sample T_c . We have done such measurements throughout much of the Brillouin zone, and have found that the gap is always maximum near M and minimum along the G-Y line, with a gap anisotropy of at least a factor 2. This anisotropy is at least an order of magnitude larger than that in any conventional superconductors (except for perhaps the heavy fermion compounds). We have compared our results with various theoretical models, and have found that all aspects of the data (the anomalously large gap anisotropy, the detailed k-space variation of the gap, and the possible existence of a node in the gap (e.g. the data at B suggests a zero gap) can be easily understood if we assume the superconducting order parameter of Bi2212 has $d_{x^2-y^2}$ symmetry. This is a very significant finding and it addresses the central issue regarding the mechanism of superconductivity in cuprates.

(3) Metal-Superconductor interfaces and the proximity effect

We were the first to apply photoemission to the study of the proximity effect [4], and we have continued these studies during the past year. In our earlier studies we chose Au and Ag as normal metal overlayers because we found these to be particularly unreactive with the cuprates. These experiments yielded a null result. A possible explanation for us not being able to detect the proximity effect induced gap is that Au has a carrier density that is two orders of magnitude higher than YBCO and Bi2212. In simple terms, this means that a very small number of superconducting quasiparticles will have to make a great number of carriers go superconducting - a very unlikely process. We are in the process of studying overlayers which should have both a low carrier density and be non-reactive with the HTSC substrate.

References

1. D. S. Dessau preprint, D. S. Dessau, Ph. D. thesis, Stanford University, 1992.
2. D. M. King, preprint, submitted to *PRL*.
3. Z. X. Shen, et al, preprint, submitted to *PRL*.
4. D. S. Dessau, et al., *Appl. Phys. Lett.* 58, 1332 (1991).

JSEP Supported Publications

1. D. S. Dessau, Z.-X. Shen, B. O. Wells, D. M. King, W. E. Spicer, A. J. Arko, L. W. Lombardo, D. B. Mitzi, and A. Kapitulnik, "Nature of the high-binding-energy dip in the low-temperature photoemission spectra of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+d}$," *Phys. Rev. B* 45, 5095 (1992).
2. B. O. Wells, Z.-X. Shen, D. S. Dessau, W. E. Spicer, D. B. Mitzi, L. Lombardo, A. Kapitulnik, A. J. Arko. "Evidence for k-dependent, in-plane anisotropy of the superconducting gap in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+d}$ " *Phys. Rev. B* 46, 11830-11834 (1992-II).
3. Z. X. Shen, "Angle-Resolved Photoemission from Normal and Superconducting State of Cuprate Superconductors, *Advances in Superconductivity IV -- Proceedings of the 4th International Symposium on Superconductivity*, October, 1991, Tokyo, Japan, page 71, edited by Hayakawa et al., Springer-Verlag 1992.
4. J. Wu, Z.-X. Shen, R. Cao, D. S. Dessau, X. Yang, P. Pianetta, D. S. Marshall, B. O. Wells, D. King, J. Terry, D. Elloway, H. R. Wendt, C. A. Brown, Heinrich Hunziker, and M. S. de Vries, "Electronic Structure of Single Crystal C_{60} " *Physica C* 197, 251 (1992).

Unit: 5

TITLE: Semiconductor Laser Structures for Optical Interconnects

PRINCIPAL INVESTIGATOR: S. S. Wong

GRADUATE STUDENT: S. Biellak

Scientific Objectives

The objective of this work is to study the interplay between geometry and physics in a semiconductor laser in order to achieve high efficiency optical emitter structures for applications in optical data communication.

Summary of Research

Typically, the width of a laser diode is constrained to be on the order of tens of microns if it is to operate with a single lateral mode (i.e., diffraction limited), as beyond this width internal self-focusing, or "filamentation", effects force multimoded operation. Consequently, single stripe AlGaAs/GaAs lasers are only capable of roughly one hundred milliwatts continuous-wave (CW) output power in a diffraction-limited spot. Commercially available multi-watt CW diode laser systems suffer from poor, multi-moded output beam quality due to the large numbers of individual lasing elements employed.

The concept of "unstable optical resonator" was developed a number of years ago and is now widely used in conventional laser devices of many different types. By using divergent rather than convergent mirrors in the laser cavity, optical rays bouncing back and forth in such a cavity will exhibit an exponential divergence outward from the axis. Rather than executing stable periodic oscillations back and forth about the cavity axis on repeated bounces as in a stable periodic focusing system, rays in an unstable cavity diverge exponentially outward from the cavity axis on repeated bounces. Despite this geometrically unstable behavior for optical rays, when the diffractive properties of optical waves are taken into account, an unstable cavity of this sort still exhibits a perfectly stable set of resonant modes or lateral cavity modes which can be used as the basis of laser oscillation. These modes extend in general across the full width of the unstable cavity, and tend to have somewhat larger output coupling than the modes in high-reflectivity stable laser resonators. There is also typically a sizable

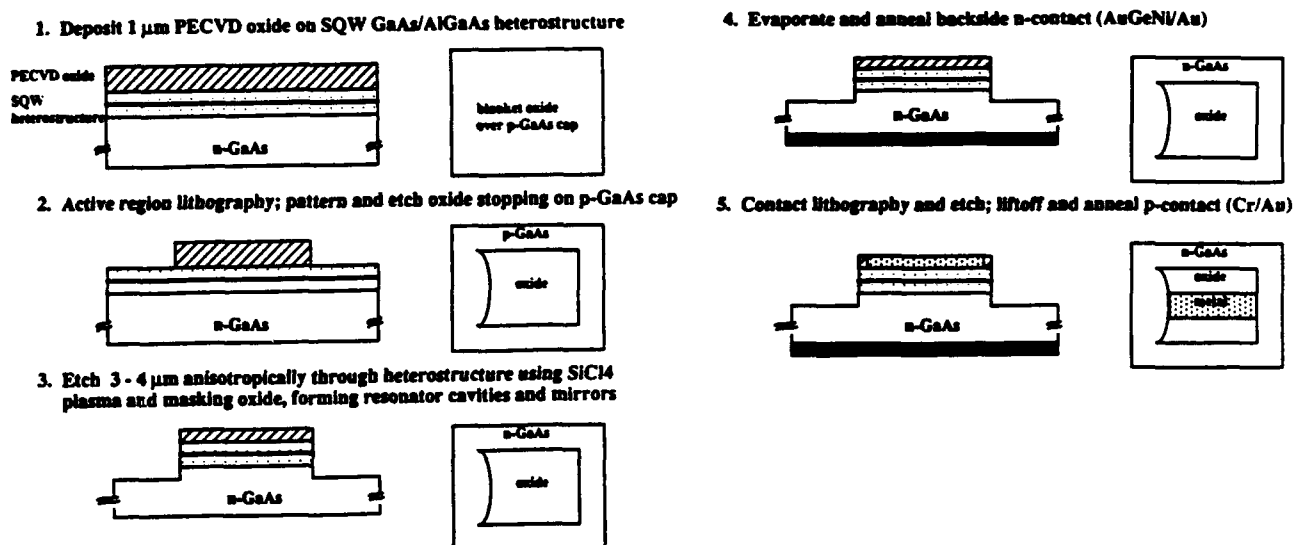


Figure 1 Process flow for unstable resonator diode laser.

discrimination between the losses of the lowest-order and the next higher-order lateral modes in an unstable resonator. As a result of this, a laser device using an unstable resonator tends to oscillate only in the single lowest-order lateral mode of the cavity. This in turn leads to a single-mode or diffraction-limited output beam. Unstable cavity modes are thus very well suited to extracting high power from the full width of a wide-stripe diode laser.

Our work investigates means of fabricating unstable optical cavities in AlGaAs/GaAs single-quantum-well heterostructures. The simplest method is to etch a curved, concave mirror facet at one end of the optical cavity, effecting beam divergence upon internal reflection of the circulating field. In the past year, we have developed a three-level process for producing unstable resonator diode lasers with such curved mirrors (Fig. 1). PECVD SiO_2 is used as the etch mask for a reactive ion etch (RIE) of the laser mirrors. Currently, both the curved and the flat cavity mirrors are etched. The RIE etch is performed using SiCl_4 gas at 20 mTorr pressure, 0.60 W/cm^2 RF power density, 200V D.C. bias, and fixed electrode temperature, producing laser facets of nearly 90° (Fig. 2). The GaAs and $\text{Al}_x\text{Ga}_{1-x}\text{As}$ layers etch at identical rates of $800\text{\AA}/\text{min}$. This low energy process helps to reduce ion damage to the facets while achieving anisotropy; the etch selectivity over the oxide mask is greater than 10:1. We have not performed detailed studies of



Figure 2 Cross-sectional SEM photograph of etched facet.

mirror surface roughness, though we have estimated it as $\lambda/5$ from high-magnification SEMs over a wide range of mirror curvatures; the facets are thus of nearly the same quality as those produced by the more complicated CAIBE and RIBE systems [Vettiger].

After the RIE etch, the AuGeNi/Au n-contact is evaporated onto the backside and annealed. The substrate is not lapped, as all devices are currently tested pulsed, and mechanical strength is required for additional processing steps. Subsequently, p-contacts are defined by an RIE etch of the remaining oxide, and liftoff and anneal of Cr/Au metallization. The p-contacts are placed 5 μm back from the mirrors to avoid pumping the mirror regions too heavily, and to minimize proximity RIE or hydrofluoric acid damage to the facets. Finally, a H_2 forming gas anneal is used to passivate the facets and increase COD levels [Ou]; however, the facets remain uncoated.



Figure 3 SEM photograph showing unstable resonator diodes with various geometries.

Recently, encouraging results have been reported for unstable resonator diode lasers fabricated using alternative processes [Srinivasan] [Tilton]. However, our approach does not require cumbersome MOCVD regrowth, nor the low-throughput focused ion-beam etching that these other groups have used. Additionally, rather than using the higher wavelength InGaAs/GaAs system, we are working at the more common 830 nm. We have successfully fabricated hundreds of unstable resonators with a variety of parameters (i.e., length, width, and geometric round-trip magnification) using our process on one 0.8 cm x 0.8 cm chip (Fig. 3). The width of the unstable cavities is defined by the current stripe; beyond this region any strongly diverging cavity rays encounter saturable loss and are attenuated.

We have tested these devices pulsed, in junction-side up configuration. As both facets are etched, our measurements do not account for the lower portion of the emitted laser beam, which is absorbed or reflected askew by the substrate resting between the laser cavity and the focusing optics. A typical diode L-I curve is shown in

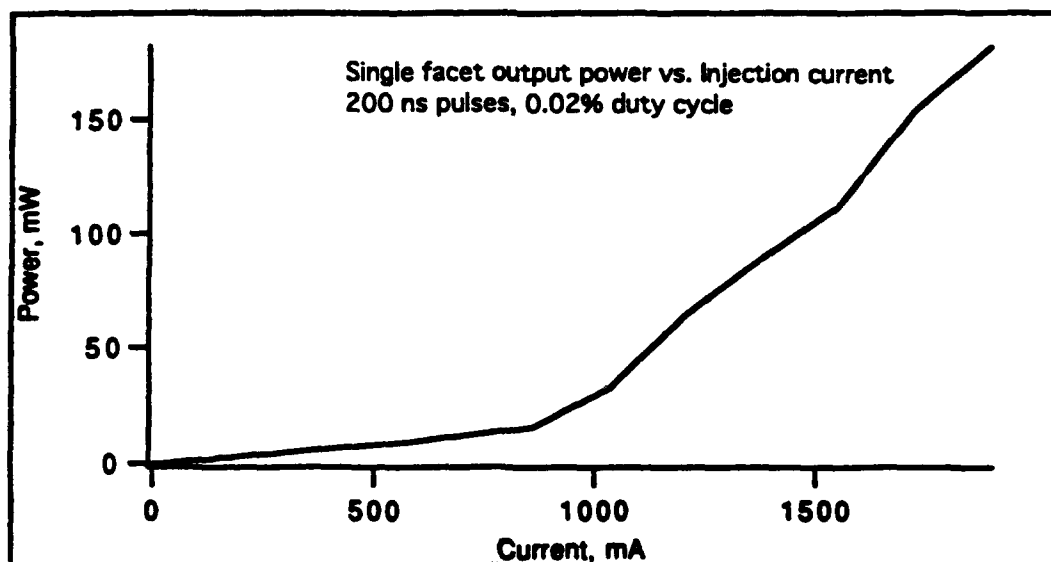


Figure 4 Typical characteristic of unstable resonator laser diode. Cavity length is $200\mu\text{m}$ and width is $100\mu\text{m}$. The mirror radius of curvature is $200\mu\text{m}$ which corresponds to a magnification of 5.8.

Fig. 4. The threshold current for this devices is somewhat higher than that of a stable cavity resonator of the same dimensions, as expected for an unstable resonator. The external slope efficiency is 0.18 W/A . The measured astigmatism is close to the value derived from geometric optics theory.

Of most interest are the values for the virtual source width and M^2 . The virtual source width is an estimate of how tightly the laser beam may be focused in the lateral dimension [Tilton]. M^2 is a measure of beam quality, normalized to that of a Gaussian; it is defined as the second moment width at the beam waist multiplied by the far-field divergence angle [Siegman]. As M^2 is invariant through lens transformations, and insensitive to near- and far-field distinctions, it is easily more measured than the usual "Times Diffraction Limited" (TDL). The diode in Fig. 4 has a virtual source width of $1 \mu\text{m}$ and a lateral M^2 of 6.0 at a current of 2x threshold. As this M^2 is considerably greater than 1, multiple lateral modes are operating in this diode. Further evidence of multimode operation has been noted in minor variations in emission intensity across the diode facet; ideally, the facet should have constant illumination. In addition, we have observed this lateral emission pattern as a function of frequency by

rotating the output beam 90° and then coupling it into a spectrometer; this allows an estimation of the number of lateral modes operating, as they have slightly different frequencies and may be visually counted if the resolution is sufficient. Finally, we are currently performing a dual-slit experiment to measure the spatial coherence across the beam width.

We are exploring means of improving the beam quality of our devices. The non-uniform near-field output patterns in our devices, though not as dramatic as typical "filamentation" observed in a wide stripe stable diode, suggest imperfections or inhomogeneities in the material, or insufficient mirror magnification (current devices have magnifications ranging from 2.5 to 7). We are now working with higher quality MOCVD material, and are actively studying the effects of increasing magnification on the devices in hopes of obtaining more diffraction-limited power. We are also considering using strained quantum well materials, instead of our current GaAs/AlGaAs structure; such systems have been shown to have much lower defect levels and may allow for higher power unstable devices [Wang].

In the future we hope to gain additional understanding of the capabilities and limitations of unstable resonator diodes; considerable improvements upon recent analyses in the literature can be made. We plan to utilize simulation code being developed at Stanford to model these devices and make additional improvements on our cavity designs. We would also like to test devices CW; it has recently been suggested local heating effects may significantly diminish the unstable operation of these devices [Yao].

References

- [Ou] S. S. Ou et al., *Appl. Phys. Lett.*, **57** (18), p. 1861, 1990.
- [Siegman] A. E. Siegman, private communication.
- [Srinivasan] S. T. Srinivasan et al, *Appl. Phys. Lett.*, **61** (11), p. 1, 1992.
- [Tilton] M. L. Tilton, et al., *IEEE J. Quantum Electron.*, **27**, p. 2098, 1991.
- [Vettiger] P. Vettiger et al, *IEEE J. Quantum Electron.*, **27**, p. 1319, 1991.
- [Wang] C. A. Wang et al, *IEEE J. Quantum Electron.*, **28**, p. 942, 1992.
- [Yao] G. Yao et al, *Optics Letters*, **17**, p 1207, 1992.

Unit: 6

TITLE: Quantum Computing

PRINCIPAL INVESTIGATOR: J. D. Plummer

GRADUATE STUDENT: B. Biegel

Scientific Objectives:

The goal of quantum computing is to use quantum effect controlled electronic devices to accomplish digital and analog computing much more efficiently, in terms of power, speed, and physical size, than in conventional electronic systems. The objective of this research project is to work towards this goal in three ways. First, our conceptual/theoretical investigations will contribute to the basic understanding of quantum computing phenomena, device operation, and computing systems. Second, and central to this research, we are developing a numerical simulation package for the accurate simulation of quantum electronic devices. Finally, we will test the simulator's predictions, and provide direction to experimental efforts, by comparing the simulated results with experimental measurements of quantum effect device operation.

Summary of Research:

During the current reporting period, our efforts were almost entirely concentrated on the main goal of this project - the development of an accurate quantum device simulation tool. In previous reports, we have described our simulation effort as a two method approach. In addition to using the two independent simulation methods to cross-check results, the methods have different strengths, which allow them to fill distinct roles in our simulation effort as summarized below:

- Transfer Matrix Method (TMM) (the traditional method) [Tsu]
- fast, efficient simulation of a wide range of structures to determine which merit more detailed study
- fine energy resolution possible
- more quickly implement, test, and perfect new simulation capabilities

- future: to model the 2nd and 3rd dimensions in multi-D simulations
- Wigner Function Method (WFM) (quantum analog of the Boltzmann transport equation) [Frensley][Jensen][Kluksdahl]
- always stable: S-C simulations always converge
- transient simulations (future: dynamic boundary conditions) are efficiently handled in a physically meaningful way
- allows additional physical effects (future: scattering) to be efficiently included in a physically meaningful way

We have made significant enhancements in the features and capabilities of both simulation tools in several areas during the current reporting period. The most important of these are the implementation of full self-consistency in both methods, the implementation of a piece-wise linear (as opposed to piece-wise constant) energy band profile for more accurate TMM simulations, and the incorporation of more accurate SiGe energy bands in both methods. These enhancements are described below.

A. Self-Consistency

Our main goal during this period was to implement self-consistency (S-C). This feature requires one to iteratively solve the quantum transport equations, which use the energy band profile (EBP) as input and give the charge distribution (CD) as output, and Poisson's equation, which does the opposite, until both agree on the EBP and CD. We have completed the implementation of S-C for our single-bias and I-V curve TMM and WFM simulators, but not yet for our WFM transient simulator (the TMM has no transient simulation capability). In general, we have found that the TMM and WFM agree substantially on S-C EBPs and CDs that they predict.

Figure 1 gives an example of both an approximate and a fully S-C EBP for a typical simulation test device, a resonant tunneling diode (RTD) [Capasso][Luryi]. The SiGe RTD used in examples for this report is based on the experimental device of [Gennser]. Note the intrinsic layer (used in most experimental RTDs), which extends well beyond both sides of the central double barrier structure. This layer produces a barrier to current flow on the emitter side that is very hard to model without full S-C. The proper treatment of such EBP features, and by implication, the implementation of full S-C, is therefore critical to getting accurate current predictions.

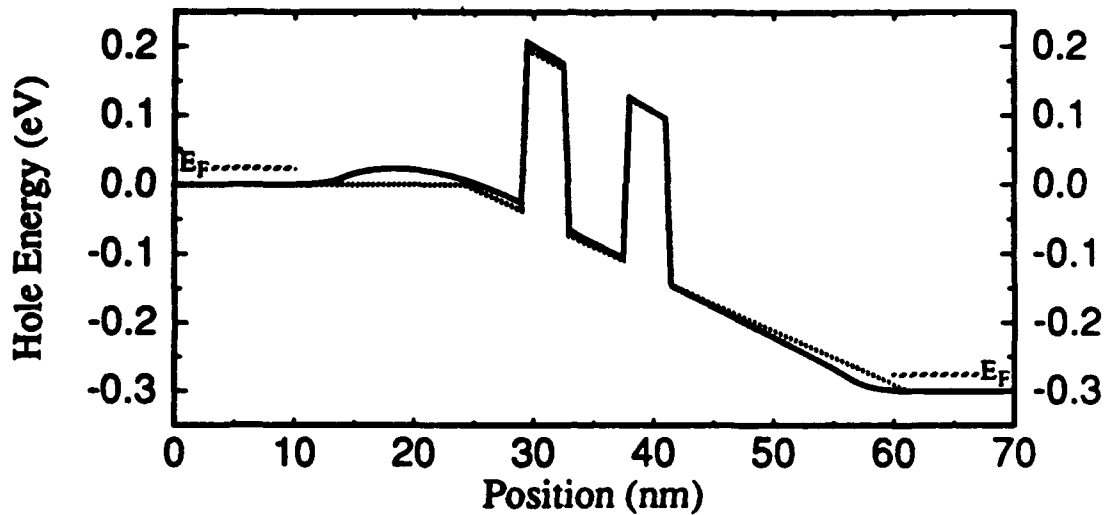


Figure 1: Linear and self-consistent RTD energy band profiles.

The iteration required to implement full S-C is computationally expensive in an already demanding numerical simulator. Thus, a major effort was made to reduce as much as possible the number of iterations necessary to reach specified convergence criterion. This requires us to provide an accurate initial guess for the EBP at each bias point. A very efficient algorithm is to simply use the EBP change from the previous step as an estimate for the change to the next step. This scheme typically gives convergence to quite rigid criteria in 1 to 4 iterations, versus 20-40 iterations obtained from other schemes.

B. Piece-wise Linear Energy Bands in the TMM

Our energy band calculation routines, S-C or otherwise, determine the value of the energy band potential only at the grid nodes. While this is sufficient input for the WFM, the TMM requires a piece-wise *continuous* potential function. In the past, we used a piece-wise constant (PWC) energy band approximation, using the values at the grid nodes. However, we were concerned that the resulting abrupt interfaces between adjacent PWC regions might be degrading simulation accuracy or creating un-real physics. To check this possibility, we implemented a piece-wise linear (PWL) energy band approximation [Tan] for the TMM. Subsequent simulations showed that the PWC and PWL approximations gave very similar results in all circumstances, assuming a reasonably fine position grid.

However, the PWL TMM simulations require the calculation of Airy functions [Abramowitz], rather than the simple sines and cosines used in PWC TMM simulations, for determining quantum wavefunctions. As a result, PWL simulations are about 10 times

slower for the same grid size. Thus, the PWL simulations are now mainly reserved to refine the results of cases of particular interest. Having the PWL capability allows us, if we choose, to use a much coarser position grid than for the PWC approximation and still get a good approximation to the energy band profile. We have not yet implemented a variable position gridding scheme for this purpose. Until and unless we see a marked advantage of the PWL over the PWC approximation, we will use the PWC option due to its speed.

One final benefit came from the development of the PWL TMM capability. Due to its increased computation time and the care necessary to implement it, we studied in great detail the calculations involved in the TMM in an attempt to make them as efficient as possible. We discovered that there are three essentially independent formulations of the main calculation, which is to determine the reflection (and transmission) coefficient of a single quantum plane wave incident on the quantum structure, in order to determine current flow. 1) The interface-based formulation (IBF) has the most straightforward derivation and implementation, but it did not take advantage of computational efficiencies that we found were possible. 2) The region-based formulation (RBF) has a more abstract derivation, but it allowed simplifications that we believed would make it the fastest way to calculate reflection coefficients. 3) The normalization-based formulation (NBF) also has a rather abstract derivation, but again allowed simplifications that made it efficient in certain cases. Further, we expected the NBF to be the fastest way to calculate the quantum wavefunction (e.g., for determination of the carrier density profile for S-C).

Table 1 shows typical TMM calculation times (in seconds) for the device current and carrier distribution using both the PWC and PWL energy band approximations. This data indicates that our expectations concerning the three TMM formulations were very accurate. The one surprise, which might have been anticipated, was that the IBF was the fastest way to calculate the wavefunction for a PWL EBP approximation. Note that the RBF is not suited to calculation of the wavefunction. Thus, we now use the RBF for calculating currents, and the faster of the NBF and IBF for calculating carrier distributions.

Table 1: Computation time comparisons for three TMM formulations

| Calculation Type | IBF | RBF | NRF | Fastest |
|------------------|-----|------|------|---------|
| Current (PWC) | 22 | 14.5 | 20.5 | RBF |
| CD (PWC) | 46 | --- | 43 | NBF |
| Current (PWL) | 130 | 124 | 210 | RBF |
| CD (PWL) | 240 | --- | 370 | IBF |

C. SiGe Energy Bands

For reasons described below, we decided that it was important to implement a more accurate model of the SiGe valence energy bands (SiGe quantum devices are usually p-type). In particular, in the past, we have assumed a single, parabolic energy band, with a carrier (hole) effective mass equal to the density of states (DOS) effective mass. In reality, there are two hole energy bands, called the light hole and heavy hole bands, each with their own effective mass. For example, in $\text{Si}_{0.5}\text{Ge}_{0.5}$, $m_{lh} = 0.10$, $m_{hh} = 0.385$, and $m_{DOS} = 0.438$, all relative to the free electron mass. With such a large variation, the choice of which mass to use in a single-mass implementation clearly has a significant impact on the simulation results.

The DOS effective mass must be used to calculate the correct boundary conditions (Fermi levels and carrier densities) for the device. Further, since the DOS, and thus the density of holes, in the heavy hole band is much greater than that in the light hole band, the heavy hole distribution versus position has more effect on the energy bands in the regions adjacent to the contacts. This is shown in Fig. 2, which is a plot of the carrier distributions for the biased SiGe RTD of Figure 1. However, because tunneling probability exponentially decreases with increasing effective mass, current through the device is mainly due to light holes [Wang][Gennser]. This is shown in Fig. 3 by a TMM simulated I-V characteristic for the SiGe RTD of the previous figures.

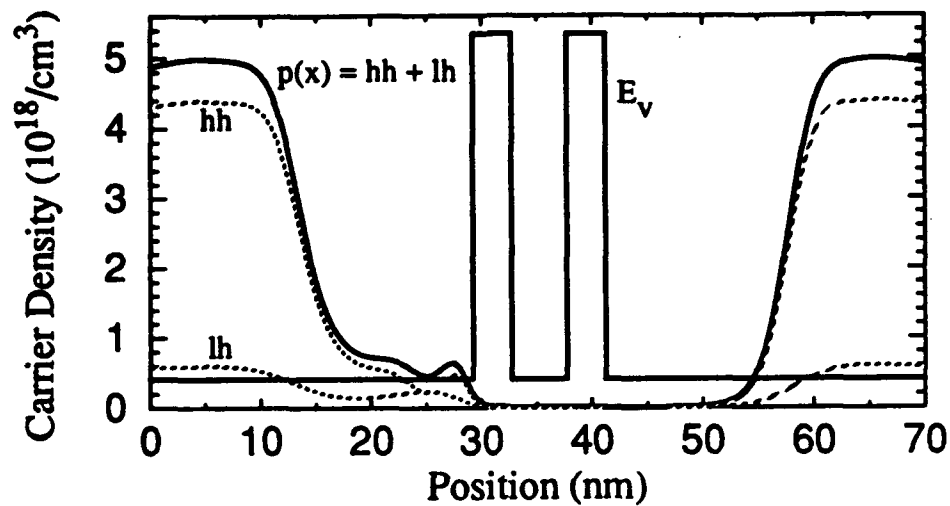


Figure 2 Carrier densities for SiGe RTD of Fig. 1.

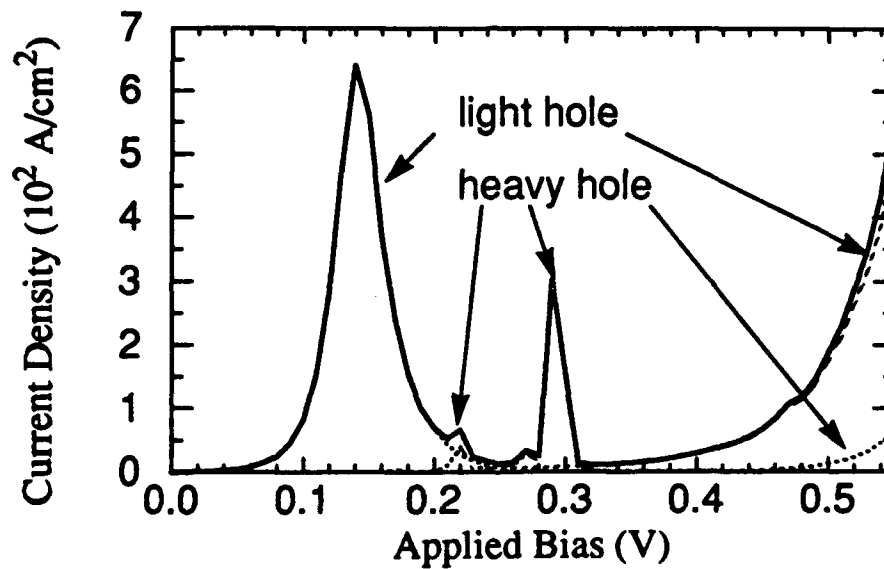


Figure 3 I-V characteristic for SiGe RTD of Fig. 1.

Future Directions:

The new features implemented in our quantum device simulation tools during the current reporting period have brought us a long way towards our ultimate goal - being able to accurately predict the operation (I-V characteristic and transient response) of real quantum effect devices. During the coming period, we intend to complete the following additional features and capabilities to make this goal a reality:

- 1) Self-consistency will be added to our transient WFM simulator.
- 2) Scattering, based on the relaxation time approximation (WFM; not possible in TMM)
- 3) Modeling the effects of external circuit elements, such as contact resistance.

The implementation of full S-C and more accurate energy bands has improved the predictive accuracy of our simulation tool such that it is now reasonable to begin comparing with experiment. As we include those effects mentioned above, the accuracy will be even better. Because of this, we will be increasing the amount of comparison between simulation and experiment, hopefully providing as much information and direction to experiment as experiment has to our simulation development efforts in the past. We will also expand our simulations to include other quantum effect controlled structures besides the RTD, including pseudo three terminal structures (by specifying the potential at an intra-device node). Other features and capabilities that will be implemented as deemed necessary or important (and as time allows) include:

- 1) A more accurate energy band and energy gap dispersion relation (effective mass) model.
- 2) Transient TMM capability.
- 3) Drifted Maxwell distribution and dynamic boundary conditions.
- 4) Dynamic energy and position gridding for improved speed and accuracy (TMM)
- 5) More correct treatment of effective mass variation in the WFM.
- 6) 2-D and 3-D simulations.

Improving simulation speed is always a major concern. We have made significant breakthroughs during the current reporting period in this regard, some of which were discussed above. We will, of course, never be satisfied in this regard.

References:

[Abramowitz] M. Abramowitz and I. A. Stegun, eds., *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, National Bureau of Standards, Washington, DC, 10th printing, Dec. 1972.

- [Capasso] F. Capasso, K. Mahammed, and A. Y. Cho, *IEEE J. Quan. Elec.*, **22**(9), 1853 (1986).
- [Frensley] W. R. Frensley. *Phys. Rev. B*, **36**(3), 1570 (1987).
- [Gennser] U. Gennser, V. P. Kesan, S. S. Iyer, T. J. Bucelot, and E. S. Yang, *J. Vac. Sci. Technol.*, **B8**(2), 210 (1990).
- [Jensen] K. L. Jensen and F. A. Buot. *IEEE Trans. Elec. Dev.*, **38**(10), 2337 (1991).
- [Kluksdahl] N. C. Kluksdahl, A. M. Krizan, and D. K. Ferry, *Computer*, **22**(8), 60 (1989).
- [Luryi] S. Luryi. *Superlat. and Microstruc.*, **5**(3), 375 (1989).
- [Tan] C. M. Tan, J. Xu, and S. Zukotynski, *J. Appl. Phys.*, **67**(6), 3011 (1990).
- [Tsu] R. Tsu and L. Esaki, *Appl. Phys. Lett.*, **22**(11), 562 (1973).
- [Wang] K. L. Wang, R. P. Karunasiri, J. Park, S. S. Rhee, and C. H. Chern, *Superlat. and Microstruc.*, **5**(2), 201 (1989).

Unit: 7

TITLE: Applications of Silicon-Germanium and Germanium Films in MOS Technologies

PRINCIPAL INVESTIGATOR: K. C. Saraswat

**GRADUATE STUDENTS: M. Cao, D. J. Connelly
and A. W. Wang**

Scientific Objective

The objectives of this work are to characterize the physical and electrical properties of polycrystalline silicon-germanium (poly-SiGe) and polycrystalline germanium (poly-Ge) films and to assess their advantages over polycrystalline silicon (poly-Si) for various applications in MOS technologies; and to fabricate confined electron heterostructures using single-crystal SiGe. In the current reporting period, we have investigated the bandgap density of states in poly-SiGe for the first time, characterized the deposition rate of poly-SiGe and poly-Ge films in a recently modified, higher quality LPCVD furnace, and demonstrated that Ge films can be deposited at temperatures as low as 300°C.

Introduction to Poly-SiGe and Poly-Ge TFTs

Thin film transistors can be used to integrate peripheral drive circuits with active-matrix liquid-crystal displays (LCDs) on large area substrates [Ohwada]. This integration greatly reduces the number of external connections and the number of external driver chips, thereby lowering cost and improving reliability. Low-temperature, short-time processing is essential in order to fabricate low-cost integrated circuits for large-area electronics. Hydrogenated amorphous silicon (α -Si:H) TFTs can be fabricated on glass substrates, but their low effective mobilities make them unsuitable for use in driver circuits for large-area displays. Poly-Si TFTs have much higher effective mobilities ($>10\text{cm}^2/\text{Vs}$); however, using conventional processing equipment, their fabrication requires higher temperatures ($>600^\circ\text{C}$) and long-time processing, making them marginally compatible with large-area glass substrates.

Properties of Poly-SiGe and Poly-Ge Thin Films

Since the melting point of SiGe is lower than that of Si, physical phenomena controlling fabrication processes such as deposition,

crystallization, grain growth, and dopant activation occur at lower temperatures for SiGe than for Si. Poly-SiGe films with Ge mole fractions up to 0.6 are completely compatible with standard VLSI fabrication processes, and can be routinely deposited in commercially available LPCVD equipment using silane (SiH_4) and germane (GeH_4) as the gaseous deposition sources [King1]. It has already been demonstrated that such films can be deposited at lower temperature (400°C), crystallized at lower temperature (550°C), and annealed for dopant activation at lower temperature (500°C) [King1] than pure poly-Si films. It has also been demonstrated that grain boundaries in such films can be passivated by hydrogen implantation [King2].

Polycrystalline Ge films and poly-SiGe films with Ge mole fractions above to 0.6 are not compatible with standard VLSI fabrication processes; these films are attacked by common cleaning solutions and can degrade in the presence of moisture. However, processing temperatures decrease further as the Ge mole fraction increases. In addition, it may be possible to grow high-quality dielectrics on these films [Hymes]. All of the properties mentioned above are conducive to a low-thermal-budget polycrystalline TFT technology. Poly-SiGe and poly-Ge films may also exhibit better electrical characteristics due to potential enhancement of carrier mobilities in SiGe [Pearsall].

Bandgap State Density of Poly-SiGe Thin Films

The electrical properties of polycrystalline semiconductors are strongly controlled by defect states in the energy band gap caused by dangling and strained bonds at the grain boundaries [Seto]. An understanding of the nature, energy distribution, density, and behavior of these defects is important for the development of polycrystalline TFTs. To improve this understanding, the gap state density was determined for hydrogenated poly- $\text{Si}_{0.8}\text{Ge}_{0.2}$ for the first time, and compared to the gap state density for poly-Si measured by the same technique.

Experimental

The experiment used previously fabricated NMOS and PMOS TFTs as described in last year's report [King3]. The field-effect conductance is thermally activated with an energy E_a , the energy difference between the Fermi level and the effective conduction-band edge (for NMOS transistors) or the effective valence band edge (for PMOS transistors). E_a can be determined for each gate voltage

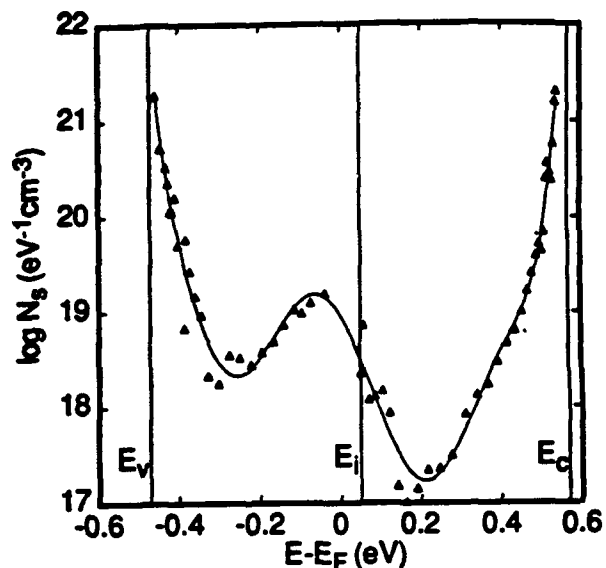


Figure 1 Poly-Si density of gap states vs. energy. The energy is referenced to the Fermi energy.

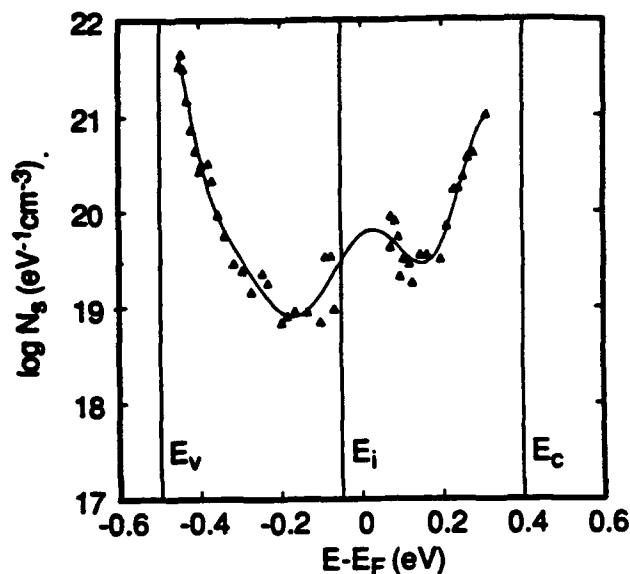


Figure 2 Poly-Si_{0.8}Ge_{0.2} density of gap states vs. energy. The energy is referenced to the Fermi energy

from Arrhenius plots of the logarithm of the field-effect conductance. Field-effect conductance measurements were carried out by sweeping NMOS gate voltages from -10V to +20V, and PMOS gate voltages from +10V to -20V, with fixed drain voltages $|V_{DS}| = 0.1V$. The measurements were performed at different temperatures (from 21°C up to 125°C) to obtain the temperature dependence of the conductance [Fortunato]. Based on this data, the energy distribution of the density of states is determined by numerically solving a one-dimensional Poisson's equation [Suzuki].

Results and Discussion

Figures 1 and 2 show the calculated density of gap states in the hydrogenated poly-Si and poly-Si_{0.8}Ge_{0.2} films. It is obvious that the Si film has a lower density of gap states than the Si_{0.8}Ge_{0.2} film (about an order of magnitude lower) and that the Si film has a larger band gap than the Si_{0.8}Ge_{0.2} film. The hydrogenated poly-Si film is p-type, with the Fermi level located 0.05eV below midgap; the hydrogenated poly-Si_{0.8}Ge_{0.2} film is n-type, with the Fermi level located 0.05eV above midgap.

Figure 1 shows that there is a "hump" below the Fermi level for poly-Si, which is caused by dangling bond states and is located at ~0.63eV below the conduction band. The poly-Si state density agrees well with optical absorption measurements [Jackson]. The

measurements were not able to clearly locate the dangling bonds states in poly-Si_{0.8}Ge_{0.2}, but an estimate of these states is shown in Fig. 2.

The larger number of gap states in the poly-SiGe films used in this work may be due to preferential Si dangling bond passivation by hydrogen [Stutzmann], incomplete dangling bond passivation, smaller grain size, oxygen contamination, or bonding strain within the grains themselves.

Improved Characterization of Poly-SiGe and Poly-Ge Thin Film Deposition

In order to improve our capability to deposit poly-SiGe, a computer-controlled LPCVD furnace has been modified specifically for the deposition of poly-SiGe has been brought on-line. This furnace exhibits substantially less leakage than our previous furnace, which should reduce contaminants such as oxygen [Ghani] which might have detrimental impact on our deposited films. In addition, the temperature stability and uniformity should be improved over the previous furnace.

Experimental

The starting substrates were silicon wafers with 100nm thermally grown oxide. SiH₄ and GeH₄ were used as the source gases with SiH₄:GeH₄ flow ratios of 1:0, 10:1, 4:1, 2:1, 1:1, 2:5, and 0:1. No hydrogen or carrier gas was used. The pressure was fixed at 100mtorr. The temperature was varied from 300°C to 650°C. The conditions were chosen to provide a reasonably high deposition rate (above 2nm/min). The films were normally grown with a target thickness of 250nm. The film thickness was determined by surface profile measurements after plasma etching. The Ge fraction was determined by X-ray diffraction to an accuracy of about 1.5 atomic percent; a selection of samples were confirmed by Rutherford backscattering spectrometry, with good agreement between the two methods.

Results and Discussion

The Arrhenius plot for the deposition rate as a function of GeH₄ fraction in the deposition gas is shown in Figure 3. The plot shows that poly-Ge can be deposited at temperatures as low as 300°C, which is comparable to the temperatures used for the plasma-assisted deposition of a-Si:H. The large increase in growth rate

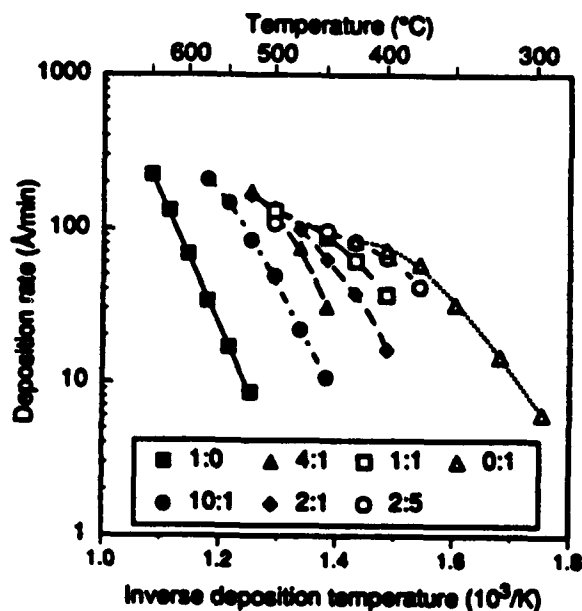


Figure 3 Arrhenius plot for the deposition rate of $\text{Si}_{1-x}\text{Ge}_x$ films on SiO_2 at 100 mtorr as a function of $\text{SiH}_4\text{:GeH}_4$ flow ratio. $\text{SiH}_4\text{:GeH}_4$ flow ratios of 1:0, 10:1, 4:1, 2:1, 1:1, 2:5 and 0:1 were used, as indicated on the plot.

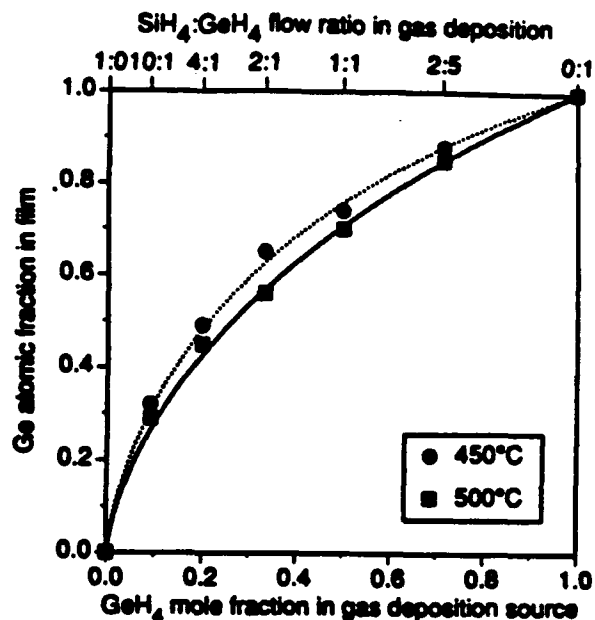


Figure 4 Ge atomic fraction in deposited film vs. GeH_4 mole fraction in deposition gas at 450°C and 500°C. The two curves are fitted using the model described in the text.

with the addition of a small fraction of GeH_4 is consistent with works on $\text{Si}_{1-x}\text{Ge}_x$ epitaxy [Garone].

The relation between GeH_4 mole fraction in the deposition gases and the Ge atomic fraction in the film is shown in Fig. 4. A model, based on the sticking probability of Si and Ge precursors, is proposed to explain this behavior. Near the surface of the wafer, the SiH_4 and GeH_4 both decompose into Si and Ge precursors, respectively. These precursors travel to the surface of the wafers and get adsorbed into adsorption sites. It is believed that there is only a small probability that the precursor will actually stick to the growing surface rather than desorb from the surface [Vossen]. The probability that the precursor will stick is affected by the type of precursor (Si or Ge); it also depends on whether the precursor lands on a Si atom or a Ge atom. The model is fitted to the isotherms for 450°C and 500°C in Fig. 4, and shows good agreement for this data.

Future Work in Poly-SiGe and Poly-Ge TFTs

During the next year, we propose to further develop SiGe thin film transistor technology, to explore Ge for use in thin film transistors, and to lay the groundwork for fabricating novel heterostructure devices on single-crystal Ge. Specific objectives of our work include:

- Further understanding of the growth mechanisms of LPCVD SiGe films. In particular, the sticking coefficients for our proposed model should be determined to verify the model.
- Further characterization of physical properties of SiGe and Ge thin films essential to TFT process optimization, such as crystallization or grain growth, thermal oxidation and nitridation, and dopant diffusion.
- Further characterization of electrical properties of SiGe and Ge thin films essential to TFT process optimization, such as n- and p-type doping, carrier mobility, and trap states at the grain boundaries.
- Further understanding of device issues in TFTs, such as leakage current and threshold voltage control.
- Improve the technology to passivate grain boundary traps in SiGe. Plasma and ion implantation techniques involving hydrogen and fluorine will be explored.
- Fabrication of thin film transistors using Si channels and SiGe sources and drains. Because of the lower processing temperature of SiGe, these structures may reduce problems with source/drain implant damage annealing.
- Development of technology for the fabrication of Ge TFTs. The development of suitable dielectrics and processing techniques compatible with Ge are especially important.
- Optimization of technology to fabricate both PMOS and NMOS TFTs at low temperatures (<500°C) in Si, SiGe, and Ge.
- Fabrication of display driver circuits using poly-SiGe films at temperatures (<500°C) compatible with low-temperature substrates used in the manufacture of flat-panel displays. This will be accomplished through collaboration with industrial partners such as Xerox.

Introduction to Confined-Electron SiGe Heterostructures

Single-crystal SiGe alloy exhibits a peak valence band energy which increases steadily with increasing Ge content. The energy of

the sixfold-degenerate X (used here to signify all six $\langle 100 \rangle$ directions) conduction band valleys is relatively insensitive to the Ge content in unstrained material. Up to approximately 80 atomic-percent Ge, these X -valleys have the lowest energy of the conduction states in the material. At higher Ge concentrations, however, the strong alloy-dependence of the eight-fold degenerate $\langle 111 \rangle$ L -valleys brings them to a lower energy.

Due to the dependence of the valence band energy on alloy content across the material spectrum, unipolar heterostructure devices built in the low-Ge regime have used holes as their carrier. n -type devices have been built, however, exploiting the strain-dependence of the conduction band minimum.

When (100) Si is deposited pseudomorphically on a thick unstrained crystalline SiGe alloy, the Si is in biaxial tension with decreased lattice spacing in the growth direction (z) and increased lattice spacing in the two orthogonal directions (x and y). The result is that electrons in the z -valleys ($[001]$ and $[00\bar{1}]$) exhibit a reduced energy relative to those in unstrained Si while the x and y valleys see an increase in the energy of their states. The advantages are two-fold. First, since the unstrained SiGe substrate has similar conduction band energies to unstrained SiGe, the Si now has a reduced conduction band energy relative to the surrounding material and electron confinement can be achieved. The second advantage is that these valleys exhibit a transverse effective mass lower than their longitudinal effective mass. Since conduction in the channel by z -valley electrons will be characterized by the lower transverse effective mass while electrons in the other 4 valleys will be subject to a mixture of the longitudinal and transverse effective masses, preferential occupation of the z valleys results in a decrease in net effective mass and a corresponding increase in mobility for appropriate carrier densities. The stress-induced electron confinement for devices in principle works for alloys from zero Ge up to 80 atomic percent Ge. However, work to date has focused on using strained Si as the channel material.

In Ge-rich material there is therefore available two mechanisms to yield electron confinement. If the unstrained starting material is (100) $\text{Si}_{0.25}\text{Ge}_{0.75}$, then application of a strained layer of pure Ge will result in a reduced conduction band energy due to the lower energy of the L -valleys (due to symmetry, the effect of the $[001]$ compression on the $\langle 111 \rangle$ L -valleys is small). Electrons will

preferentially occupy the Ge layer and confinement will be realized. Growth of a strained $\text{Si}_{0.50}\text{Ge}_{0.50}$ film on the same substrate will result in reduction of the z-valley energies relative to the unstrained material and again the electrons will preferentially occupy the grown layer.

Future Work in Confined-Electron SiGe Heterostructures

A comparison of these two techniques will be made. A strain-relief layer will be grown to grade the Ge content from that of a starting (100) Ge wafer to 75%. On this will be grown a pseudomorphic layer of either 100% or 50% Ge content. After germanium oxynitride and/or silicon dioxide insulator deposition, processing will continue using a standard low-temperature NMOS technology. Examination of channel mobilities as a function of temperature, electric field, magnetic field, and carrier concentration should yield valuable insight into the advantages offered by these materials.

References

- [Ohwada] J.-I. Ohwada *et al.*, *IEEE Trans. Elec. Dev.*, **36**, 1923 (1989).
- [King1] T.-J. King *et al.*, *IEDM Tech. Digest.*, **1990**, 253 (1990).
- [King1] T.-J. King *et al.*, *IEEE Electron Device Lett.*, **12**, 584 (1991).
- [Hymes] D.J. Hymes *et al.*, *J. Electrochem. Soc.*, **135**, 961 (1988).
- [Pearsall] T.P. Pearsall, *CRC Critical Reviews in Solid State and Materials Sciences*, **15**, 551, (1989).
- [Seto] J.Y.W. Seto, *J. Appl. Phys.*, **46**, 5247 (1975).
- [King3] T.-J. King *et al.*, *IEDM Tech. Digest.*, **1991**, 567 (1991).
- [Fortunato] G. Fortunato *et al.*, *Philos. Mag. B*, **57**, 573 (1988).
- [Suzuki] T. Suzuki *et al.*, *Jpn. J. Appl. Phys.*, **21**, L159 (1982).
- [Jackson] W.B. Jackson *et al.*, *Appl. Phys. Lett.*, **43**, 195 (1983).
- [Stutzmann] M. Stutzmann *et al.*, *J. Appl. Phys.*, **66**, 569 (1990).
- [Ghani] T. Ghani *et al.*, *Appl. Phys. Lett.*, **58**, 1317 (1991).
- [Garone] P.M. Garone *et al.*, *Appl. Phys. Lett.*, **56**, 2914 (1990).

[Vossen] J.L. Vossen *et al.*, *Thin Film Processes II*, Boston Academic Press, Boston, 293 (1991).

Publications

1. T.-J. King and K.C. Saraswat, *IEEE Electron Device Lett.*, **13** (6), 309-311 (1992).
2. M. Cao, T.-J. King, and K.C. Saraswat, *Appl. Phys. Lett.*, **61** (6), 672-674 (1992).
3. M. Cao, A.W. Wang, and K.C. Saraswat, submitted to *3rd Int'l Symp. on Process Phys. and Modeling in Semiconductor Tech.*

Unit: 8

**TITLE: Signal Processing for Wideband Digital
Portable Communications**

PRINCIPAL INVESTIGATOR: T. H. Meng

GRADUATE STUDENTS: S. Hemami and A. Hung

Scientific Objectives:

This research aims at providing portable digital video on demand through a wireless link. The three main technological issues for portable video communication to be addressed here are compression efficiency, error recovery, and low-power implementation.

Summary of Research:

- A reconstruction technique for recovering lost packets of transform-coded video data transmitted in a packet-switched network has been developed. Under packet loss rates as high as 1%, the reconstructed video maintains good visual quality with minimal flaws.
- Adaptive decoding strategies for vector-quantized video data transmitted under the influence of bursty channel errors are presented, which dramatically improve performance at bit-error rates as high as 0.1.
- Simulation and architectural design of a fixed-rate subband/lattice VQ compression scheme which demonstrates good video quality at 12 to 1 compression and superb resiliency to random bit errors have been completed.

1. Spatial and Temporal Reconstruction for Packetized Video

Packet-based transmission of digitally coded still images and video over channels exhibiting loss presents a reconstruction problem at the decoder. Compensation for information lost in transmission is required, via retransmission, redundant transmission, or reconstruction techniques using only the received information. Retransmission of lost packets is often not practical for real-time signals, and the ensuing additional network traffic caused by retransmission can aggravate the congestion that caused the initial packet loss. Redundant transmission entails transmitting extra information at the expense of an increased data rate, but if

channel capacity is limited, image quality is sacrificed to provide the extra bits for the redundant information. Thus a reconstruction scheme implemented at the decoder and requiring neither retransmission nor multiple transmission of data is desirable.

We have developed a frequency-based reconstruction technique for lost transform coefficients of known locations, applicable to both luminance and chrominance components, requiring no redundant transmission of information, and using all received information. Local image characteristics are maintained by using coefficients from neighboring blocks to interpolate lost coefficients. To ensure that the reconstructed block is smoothly connected to its neighbors, the weights are computed subject to a smoothing criterion along the block borders. The computational complexity of the proposed technique is an order of magnitude below that required in [Wang]. This technique performs well under both partial and complete loss of coefficients in a block and can also satisfactorily reconstruct multiple adjacent affected blocks. The presented technique is combined with reconstruction of temporal information to create a video reconstruction algorithm.

The video reconstruction technique was evaluated using three types of real-time video sequences: the familiar "table tennis" and "flower garden" sequences and a Jordan basketball sequence called "air". A 1% loss of packetized compressed video data was assumed. Compared with other reconstruction methods, our technique reconstructs both spatial information and low movement sequences at very low hardware cost, and the reconstructed image quality is excellent, with minimal errors occurring in fast movement and detailed sequences. When all types of data are lost simultaneously, the reconstructed sequences are quite acceptable, with only several blatant errant blocks in the 150 frame sequences.

Because of the low computational complexity of our reconstruction algorithm (about 1.2 times the number of operations required by simple decoding), using a commercially available floating point DSP chip with a 25MHz clock, the complete reconstruction algorithm should be able to compensate for up to 60% loss of all information. Under more reasonable loss rates such as the simulated 1% rate, the proposed algorithm is computationally feasible and provides video of high quality.

2. Channel Optimization of Vector Quantization

In order to transmit high-bandwidth information through the radio channel, lossy compression can be used for data such as images and video. One method of lossy compression is vector quantization (VQ) [Gersho]. In this section we describe how VQ can be optimized for noisy channels. In our model, the vector quantizer codeword indices are transmitted through a noisy channel to different receivers as shown in Fig. 1. The channel introduces errors which are represented by bit-flips in the transmitted VQ codeword indices. The bit-flips are considered bursty and independent so the model can be thought of as sending VQ codeword indices through a non-stationary binary symmetric channel. The channel is time varying and for the most time is noiseless, but it experiences occasional high bit-error corruption due to deep fades or loss of sync.

Traditionally, the approach to minimizing the effects of heavy channel bit-errors is to modify the VQ codebook generation to compensate for the effect of channel errors at a certain bit-error rate. This approach is called channel optimized vector quantization (COVQ) [Farvardin].

For wireless transmission, we expect that most of the time the bit error rate is very low to ensure good consistent quality; therefore the encoder is usually working under the noiseless channel assumption. The traditional approach in COVQ is targeted to a fixed bit-error rate, thus COVQ performs worse (sometimes significantly so) under perfect noiseless conditions. Our design goal for a decoder strategy is to dynamically optimize the decoder codebook based on estimates of the probability of bit errors observed at the decoder, while maintaining good performance during noise-free transmission. The bit-error rate estimate at the decoder can be derived from the signal power received at the decoder, the signal to noise ratio, or the error detection capability of some error correcting codes. With this adaptive method, the amount of compensation can be targeted to the requirements of each individual decoder.

Since our decoders are portable and operated by battery, the hardware requirement, especially its power consumption, used to dynamically optimize the decoder codebook is the primary concern.

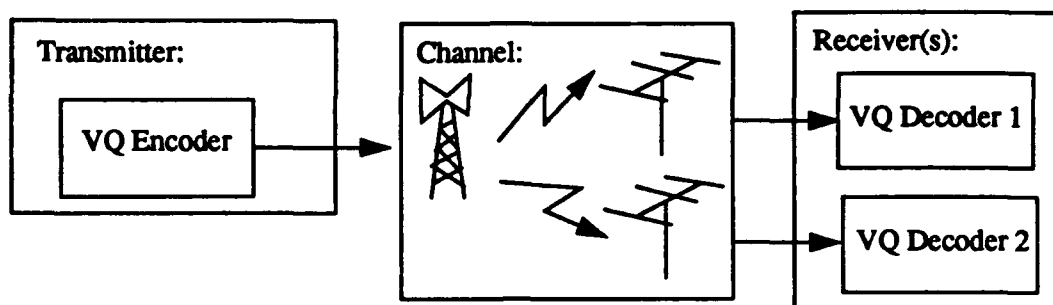


Figure 1 One transmitter sends data through a noisy channel to possibly more than one receiver.

Our VQ decoder codebook consists of a set of reproduction codewords C_i . Under the presence of channel errors, these decoder codewords can be optimized continuously with little complexity and reasonable memory overhead by *adaptive decoder optimization*. We shall first make the assumption that only single channel bit error occurs per codeword index, which significantly reduces the numerical derivation. This is a reasonable assumption for bit error probability less than 0.1. Under this assumption, the best modified decoder codeword due to channel errors, D_i , can be calculated by the following equation involving two multiplies and one addition per output pixel:

$$D_i = b_i (C_i + p_{err} A_i),$$

where the probability of a code error is p_{err} , the vector A_i is precalculated and stored in an auxiliary codebook, and the scalar factor b_i can be stored for multiple points for different, or can be efficiently calculated in real time. The memory cost of adaptive decoder optimization is one auxiliary codebook to store A_i and a set of registers per codeword index to store the scalar factors b_i at different values of p_{err} . The circuit to perform adaptive decoder optimization is shown in Fig. 2.

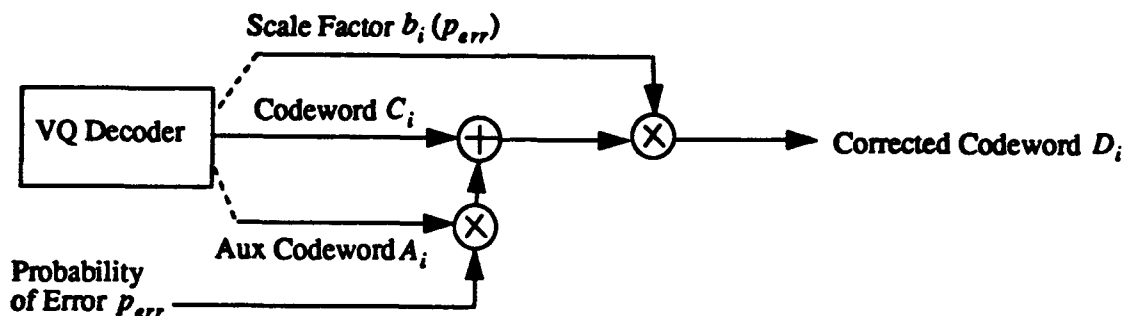


Figure 2. The hardware to perform adaptive decoder optimization.

The hardware for adaptive decoder optimization need only operate under high probability of error and the results can be cached so that the modified codewords are calculated only once.

For performance evaluation, we use a training sequence consisting of three 512x512 8-bit gray-scale images from the USC database: *couple*, *LAX*, and *man* to optimize our vector quantizer codebook. The vectors are 4x4 blocks taken from these images. The vector quantizer codebook consists of 1024 codewords, resulting in a rate of 0.625 bits per pixel.

Our simulations showed that the improvements with the channel optimized codebook (COVQ) over ordinary VQ are evident, but it has the problem of blockiness because channel optimization allocates some vectors for error protection. The blockiness will persist under noiseless conditions because the codebook was designed with a constant bit-error rate. The performance of adaptive decoder optimization, on the other hand, maintains the best possible image quality under all error conditions.

3. Indoor Wireless Digital Video

The goal of this work is to provide portable digital video transmitted over a dedicated indoor wireless link. To meet this goal, the video compression algorithm must satisfy three important criteria: error resiliency to random and bursty bit errors, good compression efficiency, and low complexity for a low power, portable design.

Our compression scheme utilizes lattice vector quantization (LVQ) together with subband coding. Lattice VQ provides several distinct advantages. First, it is a fixed-rate code, which results in hardware simplicity and prevents catastrophic error propagation. Second, because of its regular lattice structure, lattice VQ allows for simple real-time decoding and encoding. In fact, all decoding and encoding can be computed, allowing for very large codebooks without the physical memory constraint at the decoder found with regular VQ schemes which require look-up tables. Third, when optimized for the statistics of image data, LVQ provides excellent rate-distortion performance for moderate to high bit rates.

Subband coding offers additional compression by decorrelating image data via a series of low-pass and high-pass filters. Unlike the DCT, the subband approach does not introduce blockiness, and subband images are often found to be more visually pleasing. In addition, the hierarchical nature of subband filtering allows for flexibility in bit allocation, with different bit rates assigned to different bands based on information content, visual importance, and band size.

Our scheme consists of the lattice vector quantization of subband coefficients. Subband coding involves dividing each video frame into various subbands by passing the image through a series of low-pass and high-pass filters. Each stage divides the image into four subbands. The pixel values in each subband are referred to as subband coefficients. Fig. 3 shows the subband filtering for one stage. We can hierarchically decompose the image by further subdividing the lowest band. The filtering process decorrelates the image information and "compresses" the image energy into the lower bands, a natural result of the fact that most image information lies in low spatial frequencies. The higher spatial frequency information, such as edges and small details, lie in the higher subbands, which have less "energy." We achieve compression by quantizing each of the subbands at different bit rates according to amount of energy (variance) in each subband and the relative visual importance of the subbands.

A common approach to quantizing subband or DCT coefficients is the use of scalar uniform quantization followed by entropy coding. This approach, although nearly optimal in rate-distortion performance for moderate to high bit rates, results in a variable-rate code, which is very sensitive to catastrophic bit errors and

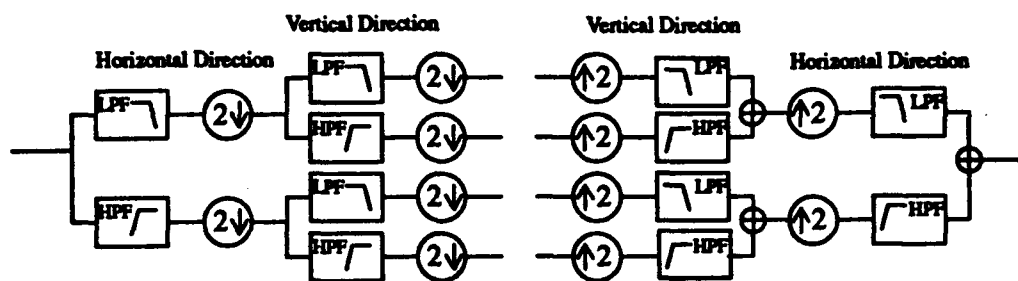


Figure 3 One stage of 2-D subband filtering.

requires significant buffering and additional hardware complexity. We instead use lattice vector quantization, which achieves comparable performance with a fixed-rate code.

In regular VQ schemes, codebook design typically involves the LBG algorithm to "train" or design locally optimal code vectors (centroids) in an N-dimensional space. In lattice VQ, regularly spaced lattice points represent the code vectors, one example of which is the regular Z_n lattice. In designing a lattice codebook with optimal rate-distortion performance, we confirmed by simulation that subband coefficients closely resemble the Laplacian distribution. For the Laplacian distribution, the "high probability" region is an N-dimensional pyramid surface, as shown in Fig. 4. Therefore to best quantize vectors of subband coefficients, we should select lattice points that fall on this pyramid surface to represent code vectors in the codebook. This pyramid VQ theoretically approaches 0.255 bits/pel from the Shannon optimal rate-distortion performance, the same performance as uniform scalar quantization followed by entropy coding.

It is important to emphasize several key advantages of this lattice VQ scheme over standard LBG-based VQ scheme. Large codebook sizes and large vector dimensions lead to lower distortion and better performance. However, with LBG-based VQ, encoding with large VQ codebooks becomes impossible in real systems. With lattice VQ, although its code vectors are not trained like LBG-based codebooks, the codebook size and vector dimension are not limited, because encoding computation complexity depends linearly on the

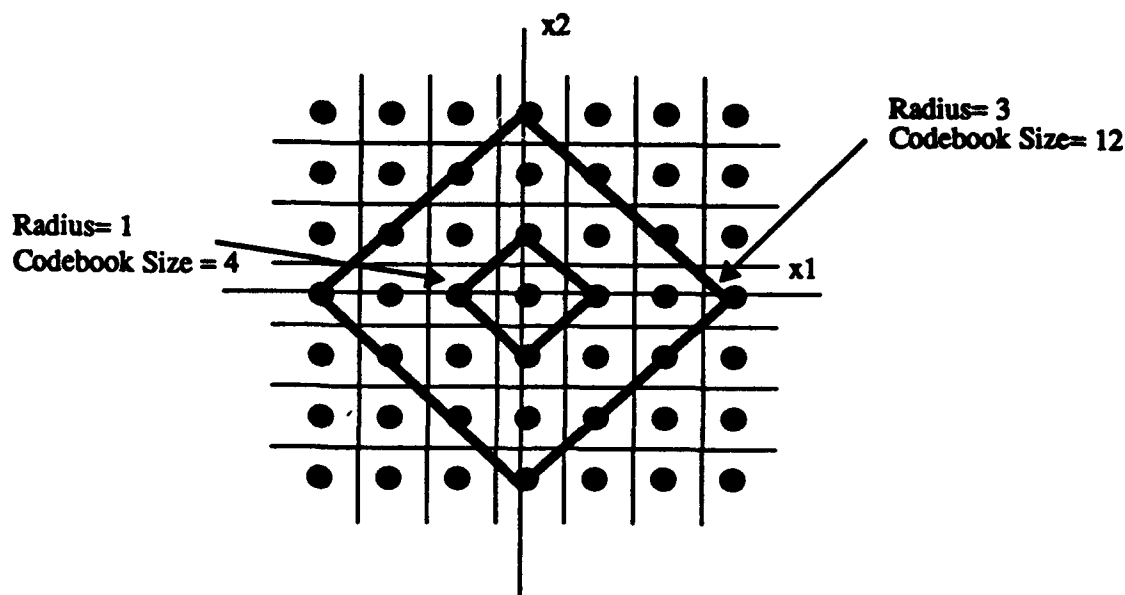


Figure 4. 2-D pyramid surface with different radii and codebook sizes.

vector size, not exponentially. This allows for real-time encoding. Furthermore, decoding can be computed (not a look-up table), eliminating large memory on the decoder chip.

The compression ratio is about 12 to 1 in our system, which compresses CIF format video (luminance at 352x240 pixels, chrominance at 176x120 pixels, frame rate of 30 frames/sec) from 30 Mbps to 2.5 Mbps. It should be noted that our algorithm is strictly an intraframe scheme, which is the reason that we make the comparison to JPEG, a variable-rate, DCT-based intraframe standard.

In comparing the subband/LVQ scheme to JPEG at 12 to 1 compression, we found that the subband/LVQ scheme results in slight fussiness and wavy artifacts around edge details. As expected, JPEG, a DCT block-based scheme, results in slightly sharper edge details, but introduces blocking artifacts in the flat regions as well as around the edge of subjects. However, both compression schemes produce good image quality. We evaluated the error resiliency of our algorithm by sending the compressed bits over a noisy channel with random bit errors. To make the comparison fair, we also added an error correcting code to JPEG, which added about 40% bandwidth overhead but improved SNR performance 2-4 dB under various BER conditions. At BER of 10^{-5} We found that the JPEG image without error correction reveals loss of synchronization

caused by bit errors, resulting in severe visible degradation in image quality. On the other hand, the JPEG image with error correction at the same data rate fixes most of the lost blocks, but reducing the bandwidth dedicated to image itself, so the overall picture quality is worse. Since our subband/lattice VQ scheme requires no error correction, close observation reveals some slight wavy artifacts in the image, but overall, the image quality is hardly degraded at this bit error rate and the image quality is maintained. At $BER = 10^{-3}$, JPEG without error correction has severe block loss and distortions, making the image unacceptably bad. With error correction, most of the errors are fixed, but again, the overall image quality suffers due the 40% rate overhead. In the subband/LVQ scheme the bit errors cause increased blurrines, but most of the image detail remains.

Since we plan to implement a LVQ decoder chip, we have performed a preliminary evaluation of the complexity and power consumption of the decoder. Evaluation of the subband decoding is discussed in the next paragraph. Decoding LVQ codeword indices involves a combination of adds, multiplies, compares, and memory fetches. Our circuit simulations showed that the power consumption of the LVQ decoder is $64mW$ at 5V supply voltage and is merely $7.3mW$ at the targeted 1.5V supply voltage.

The two main goals of the architectural design of the subband decoder are low power for portability and image size independence for support of various display devices. Low power can be achieved through lower supply voltage and increased parallelism to maintain the desired computational throughput. Also, due to large off-chip interconnect capacitance, external frame buffers and temporary data stores are eliminated. Instead, these values are kept internally, forcing the memory required to be minimized due to on-chip area constraints. To support multiple chip operation on a single image, each chip processes consecutive vertical slices of the image. A single chip should support a small portable color display of approximately 352×240 pixels per frame at 30 frames per second while multiple chips in conjunction should be capable of processing larger sized images, thus handling any size display.

The subband/lattice VQ chip set will first be modeled in C and then completely detailed in Verilog. When the final design is determined the chip will be implemented in silicon and fabricated. We also plan to further improve our algorithm's error resiliency by experimenting with selective error protection schemes

and a more clever index assignment method. We will improve the coding of the sparsely populated subbands, which have lower entropy and are poorly coded with the current scheme.

References

- [Farvardin] N. Farvardin, "A study of vector quantization for noisy Channels," *IEEE Trans. Info. Theory*, **36**, No. 4, July 1990, pp. 799-809.
- [Gersho] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic, 1992.
- [Wang] Y. Wang and Q.-F. Zhu, "Signal Loss Recovery in DCT-based Image and Video Codecs," *SPIE Conf. on Visual Communication and Image Processing*, **1605**, Boston, Nov. 1991, pp. 667-678.

Publications Supported by JSEP:

1. Sheila S. Hemami and Teresa H.-Y. Meng, "Reconstruction of Lost Transform Coefficients Using a Smoothing Criterion," to appear in *Visicom '93, Picture Coding Symposium*, March 1993.
2. Sheila S. Hemami and Teresa H.-Y. Meng, "Spatial and Temporal Video Reconstruction for Non-layered Transmission," to appear in *Visicom '93, Fifth International Workshop on Packet Video*, March 1993.
3. Andy C. Hung and Teresa H.-Y. Meng, "Adaptive Channel Optimization of Vector Quantized Data," to appear in *Data Compression Conference '93*, April 1993.

Unit: 9

TITLE: Wideband Data Transmission Techniques

PRINCIPAL INVESTIGATOR: J. M. Cioffi

**GRADUATE STUDENTS: N. Al-Dhahir, H. Bims,
H. Lou, P. Okrah and S. K. Wilson**

1 Scientific Objective

The purpose of our study has been identification of reliable and efficient methods for data transmission and detection in a mobile environment. We are particularly concerned where data rates are high and multipath fading can dominate the performance.

2 Summary of Research

The focus of our study has been in two main areas: expanded constellations in a wireless environment and maximizing channel throughput using time-domain techniques. In the first area, we are able to use a 16 QAM constellation in a simulated wireless environment using multitone modulation techniques. Wireless modulation is in general restricted to Phase-Shift-Keying Modulation due to the changing nature of the channel; we have developed a simple tracking technique that takes advantage of the multitone structure. In the second area, we investigate block-based data transmission systems, and methods to efficiently optimize the transmit filter to maximize channel throughput.

3 Wireless Multitone Modulation

Recently, the European Broadcasting Union (EBU) has developed a sound broadcasting system that employs multitone modulation techniques in a wireless environment [Plenge91]. Unlike other multitone systems where bits are assigned to each tone based on the tone's signal-to-noise ratio (SNR), the Digital Audio Broadcast (DAB) system assigns an equal number of bits to each tone. As a wireless system is subject to fades, one cannot at the receiver estimate the channel in a timely manner. The DAB system recognizes that the some tones will be subject to frequency-selective fading. It employs frequency and

time interleaving with a coding scheme to provide error protection for the tones that are subject to fades. One of the advantages of the DAB system, is that it is readily adaptable to a changing wireless environment. Also with cyclic extension, the DAB system is free of intersymbol interference as well as interblock interference. We have examined the DAB system and have developed methods to increase its data rate.

3.1 Tracking of Multitone Signals in a Fading Environment

With a time-based wideband system in a fading environment, it is necessary to track the channel variations to mitigate the effects of intersymbol interference. Tracking a channel with a large delay spread is difficult and involves many parameters. Additionally, in wireless environments, the traditional modulation scheme is a phase-shift-key (PSK) system. PSK modulation is employed to ensure efficient tracking of a fading system, where channel fading varies in time. In a frequency-based system such as the DAB multi-tone system, one only needs to track each tone. This has the advantage of a series of quick parallel calculations, over a complex channel estimation. The current DAB system employs a differential 4 PSK system, where it relies on the slowly-varying nature of the tones to estimate the differential phase of the data. We have investigated expanding the constellation size on each tone to a 16 QAM constellation, an increase from 2 bits per tone to 4 bits per tone using a simple tracking system.

We track the channel frequency component in the following manner. We set

$$\hat{w}_i = \frac{13}{16} w_{i-1} + \frac{3}{16} \frac{z_i}{\hat{x}_i}$$

where

$$z_i = w_i x_i + n_i$$

where w_i is the complex frequency component of a given tone, x_i is the channel symbol from the 16 QAM constellation, n_i is additive Gaussian white noise term, \hat{w}_i is the estimate of the channel, and \hat{x}_i is the estimate of the QAM symbol. We assume that we know the correct value of x_i for the first two symbols, and then rely upon our estimate. We estimate by x by quantizing,

$$y = z_i / \hat{w}_i$$

The data is encoded using a rotationally-invariant QAM constellation, so that if an error is made in detection of a symbol, the phase-offset will not affect the decoded bits.

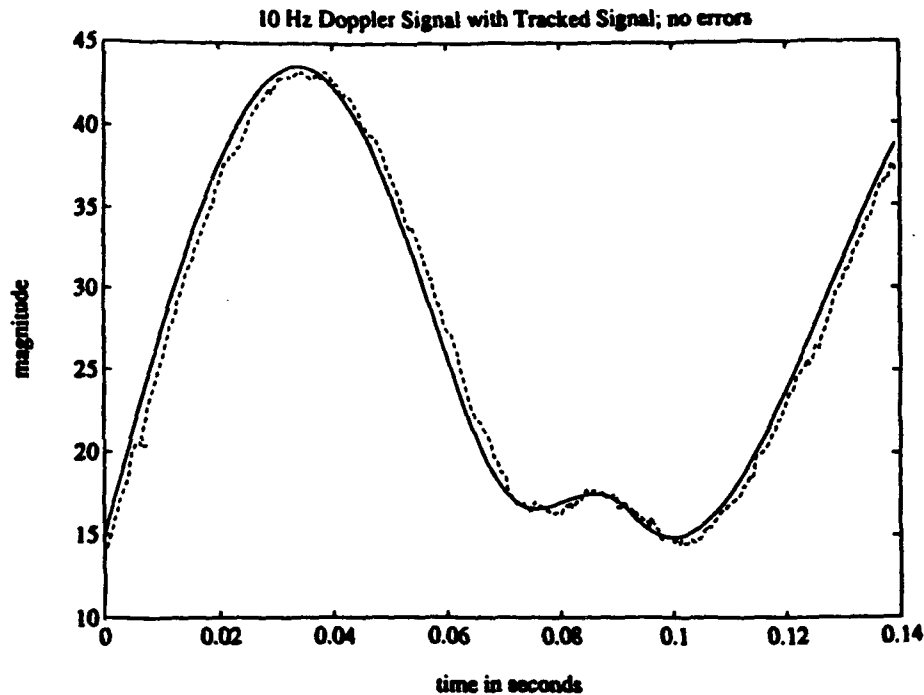


Figure 1 Tracking a 10Hz Doppler shift; no errors.

3.2 Simulation

We tracked relatively high SNR data (between 15 and 30 dB) in a fading environment. We updated each tone every 312.5 microseconds. This corresponds to a multitone block with a symbol duration of 250 microseconds and a cyclic extension of 62.5 microseconds. These parameters are the recommendation of the CCIR study group on DAB [DAB92]. Figure 1 displays the magnitude of a single tone as it varies from block to block with its estimated magnitude. We trained on the first two samples; for the remaining 444 samples in this plot, we used the estimated constellation values. In this case, no bit errors were made. Figure 2 displays another tone's magnitude as it varies from block to block, with its estimated value. In this case, two bursts of error occurred as illustrated in the plot: one between .03 and .04 seconds and another between .07 and .09 seconds. Figure 3 shows the difference between the 'true' phase and the estimated phase corresponding to the same set of tones and their estimates. Again this figure has the bursts of bit errors superimposed on the picture. We see that the error occurs when the phase is misestimated, but corrects itself when the estimated phase is $\frac{\pi}{2}$ away from the 'true' phase. A 10 Hz Doppler shift corresponds to the receiver moving at a speed of about 67 miles per hour.

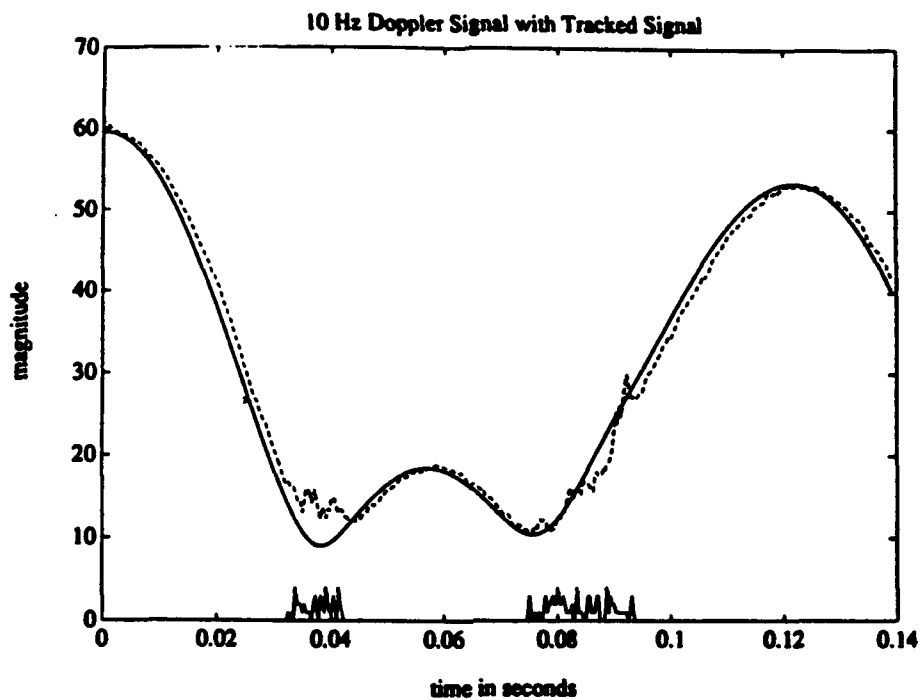


Figure 2 Tracking a 10Hz Doppler shift; with bursty error.

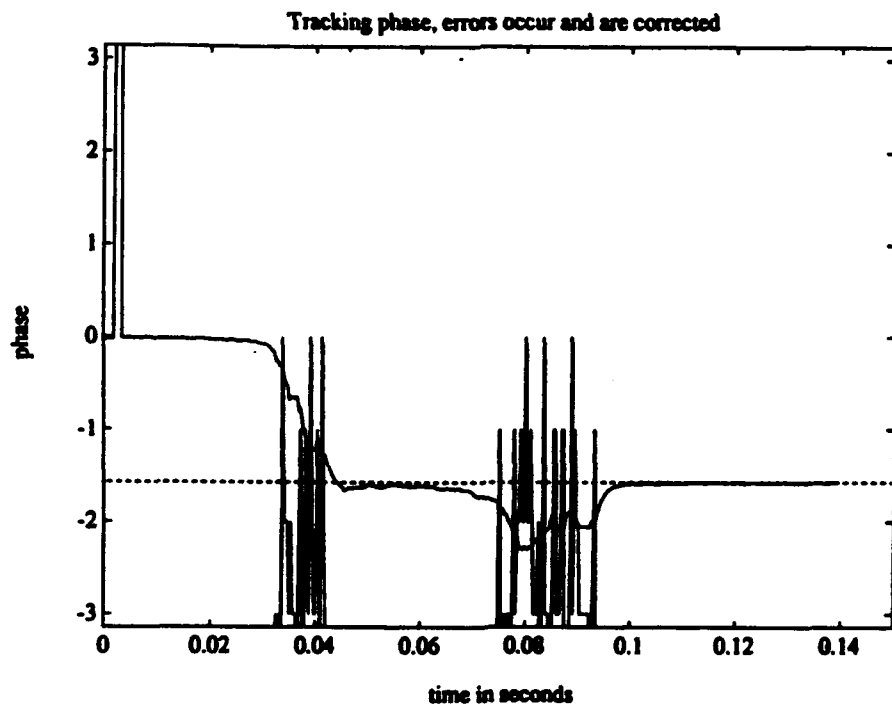


Figure 3 Phase plot corresponding to Fig. 2; with bursty error.

3.3 Applications

Though the wireless multitone modulation system is based on the European DAB, the wireless modulation scheme could be adapted to other wireless transmission schemes, such as HDTV and indoor wireless data communications. We are currently investigating where wireless multitone modulation schemes are applicable and when we could apply expanded constellations with tracking. In addition we are investigating coding schemes for error protection with the wireless multitone system.

4 Maximizing the Channel Throughput

In a typical block-based data transmission system, the received output symbols are grouped into equal-size blocks that are buffered prior to being processed to mitigate the effects of channel impairments such as additive noise and intersymbol interference. Due to the increasing demand for higher input bit rates, it is of vital importance to investigate efficient schemes for transmitting the maximum possible number of bits per input symbol, at a given symbol rate, without violating capacity limits. This goal is approached in this report by using knowledge of the channel impulse response and second order statistics of the additive Gaussian noise to introduce correlation between the transmitted symbols in a manner that maximizes the channel throughput. No prior assumption is made on the receiver structure. A finite-complexity LTI (over the duration of the block) filter is then implemented at the transmitter to achieve this optimum input correlation profile.

4.1 Statement of Problem

The input-output relationship on a discrete-time LTI dispersive channel corrupted by additive noise is given by

$$y_k = \sum_{m=0}^{\nu} h_m x_{k-m} + n_k \quad (1)$$

where $\{h_m\}_{m=0}^{\nu}$ are the channel impulse response coefficients¹ and ν is known as the channel "memory". We further assume that $\sum_{m=0}^{\nu} |h_m|^2 < \infty$.

¹The use of vector channel coefficients assumes an oversampled polyphase representation of the channel.

By considering a block of N output symbols at the receiver, we can rewrite (1) in matrix form as follows:

$$\mathbf{y} = \begin{bmatrix} h_0 & h_1 & \cdots & h_\nu & 0 & \cdots & 0 \\ 0 & h_0 & h_1 & \cdots & h_\nu & 0 & \cdots \\ \vdots & . & . & . & . & . & . \\ 0 & \cdots & 0 & h_0 & h_1 & \cdots & h_\nu \end{bmatrix} \mathbf{x} + \mathbf{n}$$

Where

$$\mathbf{y} = \begin{bmatrix} y_{k+N-1} \\ y_{k+N-2} \\ \vdots \\ y_k \end{bmatrix}; \mathbf{x} = \begin{bmatrix} x_{k+N-1} \\ x_{k+N-2} \\ \vdots \\ x_{k-\nu} \end{bmatrix}; \mathbf{n} = \begin{bmatrix} n_{k+N-1} \\ n_{k+N-2} \\ \vdots \\ n_k \end{bmatrix}$$

or more compactly,

$$\mathbf{y}_{k+N-1:k} = \mathbf{H}\mathbf{x}_{k+N-1:k-\nu} + \mathbf{n}_{k+N-1:k} \quad (2)$$

The noise symbols are assumed to be complex, zero-mean, independent and identically-distributed with power spectral density N_0 (per complex dimension). Without loss of generality the symbol rate is normalized to 1 Hertz. The input symbols are assumed complex, zero-mean, and independent of the noise with a correlation matrix (determined by the transmit filter shaping) $\mathbf{R}_{xx} \stackrel{\text{def}}{=} E[\mathbf{x}_{k+N-1:k-\nu} \mathbf{x}_{k+N-1:k-\nu}^*]$.

4.1.1 The Optimal Input Covariance

For additive Gaussian noise, the channel throughput is maximized when the input is also Gaussian [Gallager]. In this case, it is shown in [Al-Dhahir] that the throughput is given by

$$\mathcal{I}(\mathbf{X}; \mathbf{Y}) = \frac{1}{(N + \nu)} \log_2 | (\mathbf{R}_{xx}^\dagger + \frac{1}{N_0} \mathbf{H}^* \mathbf{H}) \mathbf{R}_{xx} | \quad (3)$$

where \dagger denotes the pseudo inverse.

By defining the following eigen decomposition ²

$$\mathbf{H}^* \mathbf{H} = \mathbf{U} \text{diag}(\sigma_1, \dots, \sigma_N, \underbrace{0, \dots, 0}_\nu) \mathbf{U}^*$$

it can be shown that the optimum input covariance matrix that maximizes (3) has the eigen decomposition

$$\mathbf{R}_{\text{opt.}} = \mathbf{U} \text{diag}(\delta_1, \dots, \delta_N, \underbrace{0, \dots, 0}_\nu) \mathbf{U}^* \quad (4)$$

²Note that $\mathbf{H}^* \mathbf{H}$ is a Hermitian rank N matrix.

with a corresponding optimal throughput

$$I(\mathbf{X}; \mathbf{Y})_{opt.} = \frac{1}{(N + \nu)} \sum_{i=1}^N \log_2(\max(\frac{c}{N_0} \sigma_i, 1)) \quad (5)$$

where the δ_i 's are determined from the "water pour" expression $\delta_i = \max(c - \frac{N_0}{\sigma_i}, 0) : 1 \leq i \leq N$, and the constant c is determined from the input energy constraint $\sum_{i=1}^N \delta_i = E \stackrel{def}{=} S_x(N + \nu)$.

It can be readily checked that $\delta_i \geq 0$ unless the condition number of $\mathbf{H}^* \mathbf{H}$ is very large (i.e., the channel frequency response exhibits large variations), the input SNR is low, and $N \gg \nu$. If $\delta_i < 0$ then we set $\delta_i = 0$, and the factor N in (4) is replaced by the number of positive δ_i 's, call it \bar{N} .

In this report, we shall be mainly interested in moderate-to-high input SNR levels, in which case $\bar{N} = N$. Therefore, we shall assume henceforth that $\delta_i > 0 : 1 \leq i \leq N$. In this case, (5) becomes

$$I(\mathbf{X}; \mathbf{Y})_{opt.} = \frac{1}{(N + \nu)} \sum_{i=1}^N \log_2[(SNR' + S)\sigma_i] \quad (6)$$

where $SNR' = \frac{S_x}{N_0} (\frac{N+\nu}{N}) \stackrel{def}{=} SNR \frac{(N+\nu)}{N}$, and $S \stackrel{def}{=} \frac{1}{N} \sum_{j=1}^N \frac{1}{\sigma_j}$. Moreover, $\mathbf{R}_{opt.}$ can be expressed directly in terms of \mathbf{H} and N_0 as follows :

$$\mathbf{R}_{opt.} = \mathbf{H}^* (\mathbf{H} \mathbf{H}^*)^{-1} (c \mathbf{H} \mathbf{H}^* - N_0 \mathbf{I}_N) (\mathbf{H} \mathbf{H}^*)^{-1} \mathbf{H} \quad (7)$$

This expression shows that $\mathbf{R}_{opt.}$ is non-Toeplitz, and hence the optimum finite-dimensional input covariance process is nonstationary.

4.1.2 An "almost Optimal" Quasi-Stationary Approximation

To simplify the transmit filter synthesis task that will be undertaken in Section 3, we shall approximate $\mathbf{R}_{opt.}$ by a quasi-stationary [Lev-Ari] covariance matrix $\mathbf{R}_{subopt.}$ by assuming the first N positive eigenvalues of $\mathbf{R}_{opt.}$ to be equal³, i.e.,

$$\begin{aligned} \mathbf{R}_{subopt.} &= \mathbf{U} \text{diag}(\underbrace{\frac{E}{N}, \dots, \frac{E}{N}}_N, \underbrace{0, \dots, 0}_\nu) \mathbf{U}^* \\ &= \frac{E}{N} \mathbf{H}^* (\mathbf{H} \mathbf{H}^*)^{-1} \mathbf{H} \end{aligned} \quad (8)$$

Note that $\mathbf{R}_{subopt.}$ is a projection matrix on the N -dimensional row space of \mathbf{H} .

The two terms "almost optimal" and "suboptimal" will be used interchangeably throughout this paper to describe this input correlation scheme.

³This assumption is motivated by a corresponding result in the infinite block length case which states that, for a fixed input symbol rate, using a flat input energy distribution results in negligible throughput loss from the optimal water-pour distribution, on the AWGN channel.

The throughput achieved by $\mathbf{R}_{subopt.}$ is given by:

$$\bar{I}(\mathbf{X}; \mathbf{Y})_{subopt.} = \frac{1}{(N + \nu)} \sum_{i=1}^N \log_2(1 + SN R'_i \sigma_i) \quad (9)$$

With a "flat" transmit filter (i.e., $\mathbf{R}_{xx} = \mathbf{R}_{flat} = S_x \mathbf{I}_{N+\nu}$), the channel throughput is given by:

$$\bar{I}(\mathbf{X}; \mathbf{Y})_{flat} = \frac{1}{(N + \nu)} \sum_{i=1}^N \log_2(1 + SN R \sigma_i) \quad (10)$$

Figure 4 demonstrates the relationships between $\bar{I}(\mathbf{X}; \mathbf{Y})_{opt.}$, $\bar{I}(\mathbf{X}; \mathbf{Y})_{subopt.}$, and $\bar{I}(\mathbf{X}; \mathbf{Y})_{flat}$ for a specific channel response and output block length.

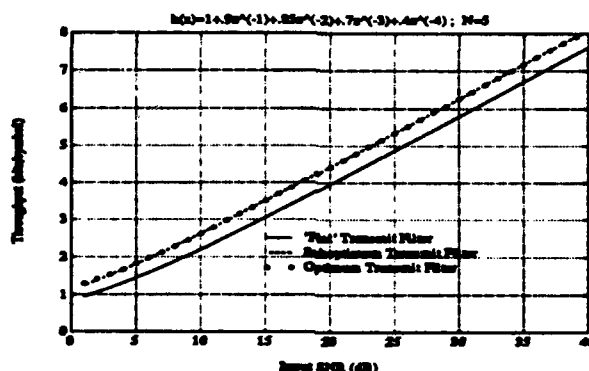


Figure 4 Effect of transmitter optimization on the channel throughput (bits/symbol).

4.2 Transmit Filter Synthesis

It is shown in [Al-Dhahir] that $\mathbf{R}_{subopt.}$, given by (8), is a "structured" matrix with a displacement rank of 2. More specifically, it is a "quasi-Toeplitz" matrix. A slightly-modified version of the classical Schur algorithm is used in [Al-Dhahir] to derive a generator [Lev-Ari] of $\mathbf{R}_{subopt.}$, call it $\mathbf{G}_{subopt.}(z)$. Once $\mathbf{G}_{subopt.}(z)$ is computed, both pole-zero and lattice realizations of the modelling filter can be derived, as shown next.

Pole-zero model:

Since $\mathbf{R}_{subopt.}$ is a "quasi-Toeplitz" matrix, it can be modelled as an $(N + \nu)$ -dimensional stationary process (with a Toeplitz covariance matrix \mathbf{T}) passed through an LTI all-zero filter $b(z)$ given by [Al-Dhahir]

$$b(z) = \mathbf{G}_{subopt.}(z) \begin{bmatrix} 1 & -1 \end{bmatrix}^t \quad (11)$$

A closed-form expression for \mathbf{T} in terms of $b(z)$ and $\mathbf{G}_{subopt.}(z)$ is given in [Al-Dhahir].

Since $R_{subopt.}$ is a singular matrix, T is also singular. It can be synthesized by an N^{th} order all-pole prediction filter $\frac{1}{a(z)}$ that can be efficiently calculated using the Levinson Algorithm [Haykin].⁴ This pole-zero model is depicted in Figure 5.



Figure 5 Block diagram of the pole zero model of $R_{subopt.}$

Lattice model:

It can be shown [Lev-Ari],[Al-Dhahir] that the N^{th} order whitening filter of $R_{subopt.}$ has the form :

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}^t \begin{matrix} N-1 \\ 0 \end{matrix} (z) \cdots \begin{matrix} 1 \\ 0 \end{matrix} (z) \begin{bmatrix} q_0 \\ -q_1 \end{bmatrix} (G_{subopt.}(z) \begin{bmatrix} q_0 \\ -q_1 \end{bmatrix})^{-1} \quad (12)$$

Hence, it consists of a cascade of the all-pole filter $k(z, q) = (G_{subopt.}(z) \begin{bmatrix} q_0 \\ -q_1 \end{bmatrix})^{-1}$, the constant-gain section $\begin{bmatrix} q_0 & -q_1 \end{bmatrix}^t$, and the N lattice sections $N-1(z) \cdots 0(z)$. This whitening filter model is depicted in Figure 6.

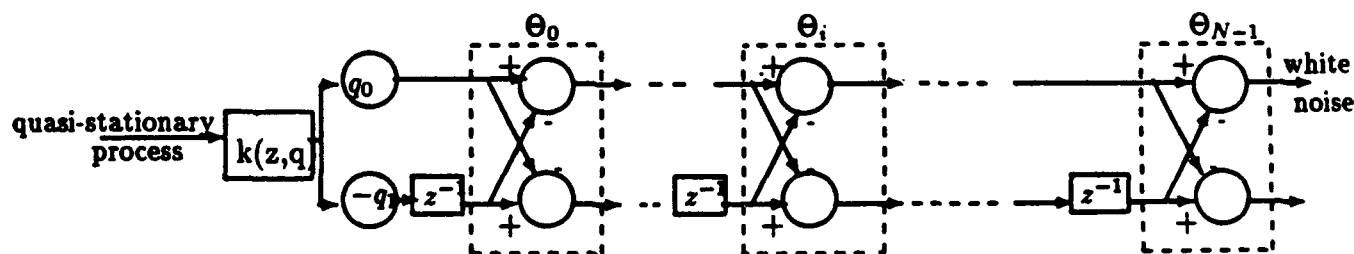


Figure 6 A lattice whitening filter for the quasi-stationary covariance $R_{subopt.}$

Once the whitening filter is constructed, the *modelling* filter is obtained by applying a standard rule of signal flow graphs : reverse the flow of the upper path, replace all gains on this path by their reciprocals, and change the signs of all inputs entering an adder along that path.

⁴Note that because of the singularity of T , the N^{th} reflection coefficient k_N will have a unity magnitude $\Rightarrow MSE_{N+1} = MSE_N(1 - |k_N|^2) = 0$. Hence, T is synthesized "perfectly".

4.3 Application to the FIR MMSE-DFE Receiver

We showed in [Al-Dhahir], that when the "almost optimal" transmit filter is implemented, the optimum output SNR of the FIR MMSE-DFE increases. More specifically, for moderate-to-high input SNR, the improvement in output SNR is given by

$$SNR_{MMSE-DFE} \approx 10 \log_{10} \left\{ \frac{(N + \nu)}{N} \right\} \text{ dB} \quad (13)$$

This analysis is extended to the optimal transmit filter case in [Al-Dhahir].

5 References

- [Al-Dhahir] N. Al-Dhahir and J. Cioffi. "Optimal Transmit Filters for Packet-Based Data Transmission on Dispersive Channels with Application to the FIR MMSE-DFE. " To be submitted to IEEE Trans. on Information Theory, January 1992.
- [DAB92] CCIR, Document 10B/TEMP, *International Symposium on Digital Audio Broadcasting* EBU Compilation , Montreux, June 8-9, 1992.
- [Gallager] R. Gallager, *Information Theory and Reliable Communication*. Wiley, 1968.
- [Haykin] S. Haykin, *Adaptive Filter Theory*. Prentice Hall, 1991. 2nd edition.
- [Lev-Ari] H. Lev-Ari, *Nonstationary Lattice-Filter Modeling*. PhD Thesis, Stanford University, 1983.
- [Plenge] G. Plenge, "DAB – A new sound broadcasting system. Status of the development – Routes to its introduction," *EBU review* no. 246. April 1991.

JSEP Supported References

- [Al-Dhahir] N. Al-Dhahir and J.M. Cioffi, "Fast Computation of Channel-Estimate Based Equalizers in Packet Data Transmission," submitted, *IEEE Transactions on Acoustics, Speech and Signal Processing*
- [Ziegler] R.A. Ziegler, N. Al-Dhahir, and J.M. Cioffi, "Nonrecursive Adaptive Decision-Feedback Equalization from Channel Estimates," submitted, *IEEE Transactions on Communications*
- [Chow] P.S. Chow, N. Al-Dhahir, J.M. Cioffi, and J. Bingham, "A Multicarrier E1-HDSL Transceiver System with Coded Modulation," submitted, *European Transactions on Telecommunications – Special Issue on Applications of Coded Modulation Techniques*
- [Al-Dhahir92] N. Al-Dhahir and J. M. Cioffi, "Fast Algorithms for the Computation of the Decision Feedback Equalizer," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, San Francisco, California, March 23-26, 1992, Vol. IV, pp. 533-536.
- [Wilson] S. K. Wilson and J. M. Cioffi, "Equalization Techniques for Direct Sequence Code-Division Multiple Access Systems in Multipath Channels," *International Symposium on Information Theory*, San Antonio, Texas, January 17-22, 1993.

Unit: 10

TITLE: Fast Computing with Neural Networks

PRINCIPAL INVESTIGATOR: T. Kailath

POSTDOCTORAL SCHOLAR: T. Constantinescu

GRADUATE STUDENTS: G. Xu and B. Hassibi

I. Scientific Objectives and Introduction

With the successful completion of K. Siu's dissertation, the neural network research focused on revision of various publications and work with him (now at UC Irvine) and Dr. V. Roychowdhury (now at Purdue) on a research monograph, tentatively titled "Discrete Neural Computation."

The main new effort was in the communications area, with an eye towards future research in wireless communications, exploiting our extensive prior research on sensor array processing. The first step was to study the consequences of signal structure of the type most commonly encountered in communications problems, namely cyclostationarity, to our previous results in this area.

II. Cyclostationarity in Sensor Array Processing

Most communications signals exhibit so-called cyclostationarity, an important signal property, which can be properly exploited to significantly improve the performance of signal processing in various applications, especially in communications. Initiated by Gardner, Franks [Gardner1] [Gardner2] and Hurd [Hurd], research activities on cyclostationarity have recently been undergoing considerable growth, which inspired the Workshop on Cyclostationary Signals sponsored jointly by the National Science Foundation, Army Research Office, and Air Force Office of Scientific Research. In this reporting period, we have focused research efforts on the exploitation of cyclostationarity for direction finding and signal estimation.

Conventional sensor array processing methods such as MUSIC and ESPRIT [Schmidt], [Roy], basically rely on the spatial properties (*e.g.* spatial delay) of the signals impinging on an array of sensors. These algorithms have the disadvantages of ignoring the signal temporal property and assuming the signals are narrow band, which may not be true in practice. In applications such as radar and sonar tracking, the signals of interest may have

rich temporal properties that can be exploited to eliminate interferences and background noise.

The cyclostationarity concept was first introduced into array signal processing by Gardner [Gardner3] and Schell *et al.* [Schell]. In their approaches, the correlation matrix estimate used in the general subspace algorithms is replaced by a cyclic cross-correlation matrix estimate (see Section II.B.). The computation of the cyclic correlation, a kind of temporal processing, helps to raise the SNR and to improve eventually the detection capability [Schell]. Another advantage of their approaches is the ability to separate signals having different cycle frequencies and thus to increase the number of detectable signals. However, the cyclic correlation methods are still based on the narrow-band assumption and thus yield larger root-mean-square error (RMSE) as bandwidth increases. In this report, we presented a new type of algorithm [Xu] that not only exploits the underlying signal spectral correlation properties but is also *asymptotically exact* for both broad band and narrow band signals.

II.A. Cyclostationarity and Spectral Correlation

It will be useful to begin our presentation with a brief discussion of some basic concepts of cyclostationarity signals.

A nonstationary process $x(t)$ is said to exhibit cyclostationarity if there exists a cyclic frequency α for which the *cyclic autocorrelation function*

$$R_x^\alpha(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} R_x \left(t + \frac{\tau}{2}, t - \frac{\tau}{2} \right) e^{-j2\pi\alpha t} dt \quad (1)$$

is not identically zero, where $R_x(t, u) = E \left[x(t) x^*(u) \right]$ and the asterisk denotes complex conjugation. Note that for a stationary process, $R_x^\alpha(\tau)$ will be zero for all non-zero α .

The *cyclic spectrum* is defined as

$$S_x^\alpha(f) = \lim_{f \rightarrow 0} \lim_{t \rightarrow \infty} \frac{1}{t} \int_{-t/2}^{t/2} f E \left\{ X_{1/f} \left(t, f + \frac{\alpha}{2} \right) X_{1/f}^* \left(t, f - \frac{\alpha}{2} \right) \right\} dt \quad (2)$$

where

$$X_{1/f}(t, v) = \int_{t-\frac{1}{2f}}^{t+\frac{1}{2f}} x(u) e^{-j2\pi uv} du.$$

Moreover, as shown in [Gardner4],

$$S_x^\alpha(f) = \int_{-\infty}^{\infty} R_x^\alpha(\tau) e^{-j2\pi f\tau} d\tau.$$

Therefore, if a process $x(\cdot)$ exhibits cyclostationarity with cycle frequency α in the time domain, then in the frequency domain it exhibits spectral correlation at a shift α .

II.B. General Data Model

Here, we present a general data model that holds for both narrow-band and wide-band signals. To simplify the following presentation, we assume that all the sensors of an array have the same linear time-invariant response and that the sensors are uniformly spaced on a line. Then the data received by the array at the time instant t can be expressed as

$$\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_m(t)]^T, \quad (3)$$

$$x_i(t) = \sum_{k=1}^d s_k(t + (i-1)D \sin \theta_k) + n_i(t), \quad i = 1, 2, \dots, m, \quad (4)$$

where $s_k(\cdot)$ and θ_k are the received signal and the Direction-of-Arrival (DOA) arising from the k -th source; c is the wave speed and D represents the separation of the neighboring sensors. The noises, $n_i(\cdot)$ at the different sensors are assumed to be jointly stationary (so that they have identically zero cyclic self and cross correlation). It is assumed that d sources are mutually *cyclically uncorrelated* at certain cycle frequency α and there are $d_\alpha \leq d$ sources sharing the same cycle frequency α .

III. New Signal Subspace Fitting Algorithms

Let us first note a simple but important property of cyclic correlation, which is given in Chapter 12 of [Gardner5] and in Chapter 11 of [Gardner6] for time series (e.g., sample paths of stochastic processes).

Lemma 1 *If $x(\cdot)$ is a cyclostationary process with (self) cyclic correlation function $R_x^\alpha(\tau)$, and $y(t) = x(t + T)$, then*

$$R_y^\alpha(\tau) = R_x^\alpha(\tau) e^{j2\pi\alpha T}$$

Now, let us start from the general data model (3)-(4). The cyclic correlation of the signal received by each sensor is

$$\begin{aligned} R_{x_i}^\alpha(\tau) &= \langle x_i(t + \tau/2) x_i^*(t - \tau/2) e^{-j2\pi\alpha t} \rangle \\ &= \langle \sum_{k=1}^d s_k(t + \tau/2 + (i-1)D \sin \theta_k/c) \sum_{l=1}^d s_l^*(t - \tau/2 + (i-1)D \sin \theta_l/c) e^{-j2\pi\alpha t} \rangle. \end{aligned}$$

Since the $s_i(\cdot)$ are mutually not cyclically correlated, and since only d_α sources are self cyclically-correlated with cycle frequency α , the double summation reduces to a single sum,

$$R_{x_i}^\alpha(\tau) = \sum_{k=1}^d \langle s_k \left(t + \frac{\tau}{2} + (i-1) \frac{D \sin(\theta_k)}{c} \right) s_k^* \left(t - \frac{\tau}{2} + (i-1) \frac{D \sin(\theta_k)}{c} \right) e^{-j2\pi\alpha t} \rangle$$

$$= \sum_{k=1}^{d_\alpha} R_{s_k}^\alpha(\tau) e^{j2\pi\alpha(i-1)D \sin \theta_k/c}. \quad (5)$$

Note that by assumption, the additive noises $n_i(\cdot)$ do not have cycle frequency α and therefore make no contribution to $R_{s_i}^\alpha(\tau)$. If we collect these functions in a vector

$$\mathbf{R}_s^\alpha(\tau) = [R_{s_1}^\alpha(\tau), R_{s_2}^\alpha(\tau), \dots, R_{s_m}^\alpha(\tau)]^T,$$

we see that

$$\mathbf{R}_s^\alpha(\tau) = \mathbf{A}(\alpha) \mathbf{R}_s^\alpha(\tau), \quad (6)$$

where $\mathbf{A}(\alpha) = [\mathbf{a}_1(\alpha), \dots, \mathbf{a}_{d_\alpha}(\alpha)]$, $\mathbf{a}_k(\alpha) = [1, e^{j2\pi\alpha D \sin \theta_k/c}, \dots, e^{j2\pi\alpha(m-1)D \sin \theta_k/c}]^T$ and $\mathbf{R}_s^\alpha(\tau) = [R_{s_1}^\alpha(\tau), R_{s_2}^\alpha(\tau), \dots, R_{s_{d_\alpha}}^\alpha(\tau)]^T$.

Thus, we have the nice result that whether the original data are broad-band or narrow-band, when evaluated individually, the cyclic correlations of the data *exactly* obey a narrow-band data model with cycle frequency α as a "center frequency." We can then estimate the source DOAs by applying conventional narrow-band algorithms [Schmidt], [Roy]. For example we can proceed as follows.

THE ALGORITHM

Step 1: For signal $x_i(\cdot)$ received the i -th sensor, and for each cycle frequency α (estimated or known) estimate the cyclic correlation $R_{s_i}^\alpha(\tau)$, where $\tau = 0, T_s, \dots, NT_s$; $i = 1, 2, \dots, m$; and T_s is the sampling period.

Step 2: For each α , form a pseudo-data matrix

$$\mathbf{X}(\alpha) = \begin{pmatrix} R_{s_1}^\alpha(0) & R_{s_1}^\alpha(T_s) & \dots & R_{s_1}^\alpha(NT_s) \\ R_{s_2}^\alpha(0) & R_{s_2}^\alpha(T_s) & \dots & R_{s_2}^\alpha(NT_s) \\ \vdots & \vdots & & \vdots \\ R_{s_m}^\alpha(0) & R_{s_m}^\alpha(T_s) & \dots & R_{s_m}^\alpha(NT_s) \end{pmatrix}, \quad (7)$$

where N is determined such that for $\tau \leq NT_s$, $R_{s_i}^\alpha(\tau)$ are nonzero and significantly varying.

Step 3: For each α , pick your favorite SSF (Signal Subspace Fitting) algorithm *e.g.*, MUSIC or ESPRIT, to estimate the DOAs of signals of interest based on this pseudo-data matrix $\mathbf{X}(\alpha)$ or its sample covariance matrix $\mathbf{X}(\alpha) \cdot \mathbf{X}(\alpha)^H$.

Extensive computer simulations have been conducted in several different scenarios and the results have demonstrated the new method performed considerably better than the earlier methods. It is also important to note that the proposed method does not require interconnections between different sensors and eliminates many implementation problems such as coupling and cross talk.

IV. Summary

Most existing approaches, see *e.g.*, those in [Haykin] and [Orfanidis], to the direction of arrival problem ignore the temporal characteristics of the signals. Following Gardner [Gardner3] and Schell *et al.* [Schell], we have shown that cyclostationarity of the signals, a situation common in many communications problems, can be exploited to considerable advantages. Our new method effectively combines the temporal and spatial properties, ease of implementation, and applicability to both narrow-band and broad-band signals.

References

- [Gardner1] W.A. Gardner, *Representation and Estimation of Cyclostationary Processes*, PhD thesis, University of Massachusetts, Amherst, MA 1972.
- [Gardner2] W.A. Gardner, "Characterization of Cyclostationary Random Signal Processes," *IEEE Trans. on Info. Theory*, IT-21:4-14, 1975.
- [Gardner3] W.A. Gardner, "Simplification of MUSIC and ESPRIT by Exploitation of Cyclostationarity," *Proc. IEEE*, 76(7):845-847, July 1988.
- [Gardner4] W.A. Gardner, "Measurement of Spectral Correlation," *IEEE Trans. ASSP*, ASSP-34(5):1111-1123, May 1986.
- [Gardner5] .A. Gardner, *Introduction to Random Processes with Application to Signals and Systems*, Macmillian, New York, 1985.
- [Gardner6] W.A. Gardner, *Statistical Spectral Analysis: A Nonprobabilistic Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1987.
- [Haykin] S. Haykin, *Array Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1984.

- [Hurd] H.L. Hurd, *An Investigation of Periodically Correlated Stochastic Processes*, PhD thesis, Duke University, Durham, North Carolina, 1969.
- [Orfanidis] S.J. Orfanidis, *Optimal Signal Processing – An Introduction*, Macmillan Publishing Co., New York, 1985.
- [Roy] R.H. Roy, *ESPRIT, Estimation of Signal Parameters via Rotational Invariance Techniques*, PhD thesis, Stanford University, Stanford, CA, August 1987.
- [Schell] S.V. Schell, R.A. Calabretta, W.A. Gardner, and B.G. Agee, "Cyclic MUSIC Algorithms for Signal-Selective Direction Finding," *Proc. ICASSP '89 Conf.*, vol. 4, pp. 2278-2281, Glasgow, Scotland, May 1989.
- [Schmidt] R.O. Schmidt, *A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation*, PhD thesis, Stanford University, Stanford, CA, November 1981.
- [Xu] G. Xu and T. Kailath, "Direction-of-Arrival Estimation Via Exploitation of Cyclostationarity – A Combination of Temporal and Spatial Processing," *IEEE Trans. Signal Processing*, SP-40(7):1775-1786, July 1992.

JSEP SUPPORTED PUBLICATIONS

Journal Papers

1. B. Ottersten, M. Viberg, and T. Kailath, "Analysis of Subspace Fitting and ML Techniques for Parameter Estimation from Sensor Array Processing," *IEEE Trans. Signal Processing*, 40(3):590-600, March 1992.
2. G. Xu and T. Kailath, "Direction-of-Arrival Estimation Via Exploitation of Cyclostationarity – A Combination of Temporal and Spatial Processing," *IEEE Trans. Signal Processing*, 40(7):1775-1786, July 1992.

Published Conference Papers

3. G. Xu, R.H. Roy, and T. Kailath, "Detection of Source Number with Centro-Symmetric Arrays," *Proc. Int'l. Conf. ASSP*, San Francisco, CA, April 1992.
4. K.Y. Siu, V. Roychowdhury, and T. Kailath, "On the Complexity of Neural Networks with Sigmoidal Units," *Proc. IEEE Workshop on Neural Networks for Signal Processing*, Copenhagen, Denmark, August 1992.
5. G. Xu, A. Paulraj, Y. Cho, and T. Kailath, "Maximum Likelihood Detection of Co-channel Communication Signals via Exploitation of Spatial Diversity," *Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove, CA, November 1992.
6. L. Tong, G. Xu, and T. Kailath, "Blind Identification and Equalization via Exploitation of Cyclostationarity," *Workshop on Cyclostationary Signals*, Yountville, CA, August 1992.

PRESENTATIONS

Conference Presentations

1. Fourth Annual Rockwell International Control/Signal Processing Conference, Anaheim, CA, January 1992.
2. SIAM Conference on Linear Algebra for Signal Processing Workshop, Minneapolis, MN, February 1992.
3. International Conference on Acoustics, Speech and Signal Processing, San Francisco, CA, March 1992.

4. Applied Linear Algebra for Signal Processing Workshop, Minneapolis, MN, April 1992.
5. Workshop on Cyclostationary Signals, Yountville, CA, August 1992.
6. International Symposium on Scientific Computing and Mathematical Modelling, Bangalore, India, December 1992.

Seminar Presentations

7. University of Maryland, Systems Research Center Colloquium, February 1992.
8. ETH-Zentrum, Automatic Control Laboratory, Zurich, Switzerland, May 1992.
9. University of Strathclyde, Industrial Control Centre, Glasgow, Scotland, May 1992.
10. National University of Singapore, July 1992.
11. Indian Institute of Science, Bangalore, India, July 1992.
12. Hong Kong University of Science and Technology, August 1992.
13. Massachusetts Institute of Technology, Boston, MA, September 1992.
14. Berkeley Engineering Systems Colloquium, Berkeley, CA, November 1992.

Meeting Presentations

15. Hitachi Technical Advisory Board, March 1992.
16. Defense Sciences Research Council Workshop, La Jolla, CA, July 1992.
17. Hitachi Sixth Research and Development Advisory Panel Meeting, Tarrytown, NY, September 1992.
18. 1992 NAE Annual Meeting, Washington, DC, September 1992.

**TITLE: Research in Image Compression,
Data Compression and
Network Information Flow**

PRINCIPAL INVESTIGATOR: T. Cover

**GRADUATE STUDENTS: A. Agashe, V. Castelli, E. Erkip,
A. Lapidoth, A. Nobel, E. Ordentlich,
S. Tai, M. Trott and R. Wesel**

1 Scientific Objectives

The objective of our work is to apply the techniques of information theory to the problems of image compression and image transmission, distributed data compression and storage, and network information flow. During the past year, this work resulted in two supported papers, five conference presentations, and three Ph.D. theses.

2 Summary of Research

Several research projects are under way. We have begun an experiment to determine the limits of lossy data compression of images. We plan to empirically bound the minimum rate at which an image can be encoded without significantly distorting it. Our bound should show that substantial improvements in lossy image compression are possible.

We have continued the work started during the previous year on the fault tolerant distributed storage of information among several sites which are failure prone. During this report period Jim Roche published his PhD Thesis [Roche a] on this topic. His work proves constructively that an apparently optimistic bound on the amount of information that can be reliably stored turns out to be *achievable*. These results were presented [Roche c] at the recent 1993 International Symposium on Information Theory.

Progress in image transmission in many transmission environments is closely related to the development of effective modulation schemes. The work of Mitchell Trott in developing a theory for linear systems over groups will be crucial to advances in trellis coded modulation.

In [Nobel], Andrew Nobel extended the classical law of large numbers from a single function to an entire class of functions. This powerful result will allow a more rigorous understanding of consistency issues in the study of learning for generalized feedforward networks.

In [Castelli and Cover], we report new results which quantify the relative value of labeled and unlabeled samples for the classification of a sample based on the previous

labeled and unlabeled samples in pattern recognition. This work provides insight into how to efficiently use unclassified and classified data to train a classification system.

Also, we are continuing to explore the role of feedback in communication, and we are refining our understanding of the entropy of random trajectories.

5 Detailed Research Descriptions

3.1 Image Compression

We have begun an experiment to determine the limits of lossy data compression of images. Specifically, we are interested in finding an accurate estimate of the minimal rate at which an image can be coded without incurring significant perceptible distortion. We do this by first having one experimental subject simplify a given image without significantly distorting it, and then have another subject predict the simplified image, pixel by pixel, as accurately as possible.

The accuracy of the second subject's predictions can be quantified to yield an estimate of the entropy of the simplified image. This two-stage process — simplification followed by noiseless data compression — can be mechanized and can be proven to be optimal in a rate distortion sense. Thus not only will our estimate of the minimum achievable rate with little or no perceptible distortion be useful as a benchmark to researchers in the field, but our experimental framework itself may pioneer a new algorithm for data compression. Our bound on the achievable compression should show that substantial improvement over current schemes is possible.

As far as details of the experiment are concerned, we plan to use human beings to 'cartoon' or simplify the existing image into one which is very predictable by another human being. This simplified image must also be psychologically nearby in that it is hard for a human observer to distinguish it from the original. This distortion measure is not quantitative, but it is clear that when the two images are almost indistinguishable, the lossy compression scheme is working well.

We will ask a person who has not previously seen the original or simplified image to predict at each pixel what shades are likely to appear in the next pixel given the pixels that have been seen so far. This person will then bet in these proportions on what shade will be next. It can be proved that if the subject achieves a wealth of 2^k starting from a one unit bet on the image, then k bits can be saved in the descriptive complexity of that image. Thus, an n -pixel image will require $n - k$ bits to describe. This is essentially the method used in [Cover and King] to find the entropy of English text. (It was found that a noiseless data compression of English text of a factor of 4:1 could be achieved, and it was argued

that no further reduction was possible.)

3.2 Distributed Information Storage

Using techniques from Galois field theory and network flow theory, in [Roche a] and [Roche b], we show how to distribute information among several different sites with a minimum of redundancy so that no data is lost even if several sites become inaccessible.

A simple example of the powerful results of this work involves a sequence of $2t$ bits of which t bits are distributed to each of K sites in such a way that if all but two of the sites fail, the original sequence can be recovered. How many different sites can be used? Clearly there are at least three since we could store the first half of the sequence at the first site, the second half at the second site, and the exclusive-or of the first half and the second half at the third site. Using combinatorics, Roche shows that $K \leq 2^t + 1$. Using Galois field theory, Roche surprisingly shows this bound to be *achievable*.

Having found theoretical limits on the amount of information that can be reliably stored in various networks, attention is focused on the configuration of disks that is widely used in practice. Roche demonstrates how to reduce the time spent updating parity-check disks when data disks are continually being modified.

3.3 Linear Systems over Groups

In [Trott], a theory of linear systems over groups is developed that parallels the classical theory of discrete time linear systems over fields. This theory is an extension of the reformulation of classical linear systems over fields found in [Willems]. Using only the elementary tools of group theory, we rederive such concepts as minimal state spaces, canonical realizations and minimal state observers. As fields can be viewed as a structured subclass of groups, our results apply to ordinary linear systems as a special case. We expect that much of linear system theory can be shown to be essentially group-theoretic in nature.

Willems' viewpoint is particularly relevant to coding theory, where the collection of codewords (or code sequences) is more important than the input-output mapping from information sequences to code sequences. A set of codewords does not usually contain an explicit description of a system that generates the set. Instead, descriptions of the codewords, embodied as input-state-output realizations, are constructed to satisfy implementation goals such as minimizing complexity.

In the context of this framework, we show that any linear system C over a group G has a minimal state space at each time that is unique up to isomorphism. We construct

a canonical feedforward input-state-output realization for C that incorporates this minimal state space. We also construct a minimal feedforward observer whose state tracks the state of the input-state-output realization. The structure of the observer guarantees that a finite number of errors in the observed sequence yield only a finite number of errors in the recovered state sequence. Such an observer is readily converted into an inverter, or, by adding feedback, into a second canonical type of minimal input-state-output realization.

A primary application of this work is the representation of trellis codes as linear systems over groups of Euclidean space isometries. A trellis code with such a representation satisfies strong symmetry and regularity properties that are difficult to establish directly. Using the theory of linear systems over groups, we have found a simple geometric description of the "nonlinear" V.32 trellis code used in most 9600 bit/s commercial modems.

Our reformulation of the results of linear systems over fields often yields simple, illuminating proofs that rely only on elementary tools of group theory. In future work, we plan to recast additional results from linear system theory into a group setting. Also, by building on [Forney], and on the analysis and synthesis results of [Loeliger and Mittelholzer], we hope to derive a constructive theory of trellis codes that explicitly relates the algebraic structure of the trellis code and the geometric properties that determine its performance.

3.4 Uniform Laws of Averages

In [Nobel], the classical law of large numbers is extended from a single real valued function to a collection of such functions. This extension provides a methodology for studying the consistency of inductive procedures for statistical estimation problems, including pattern recognition, vector quantization, and the training of neural networks. Nobel uses closure properties that hold when the underlying observations are independent and identically distributed to extend uniform laws of averages to empirical quantities such as maxima or minima that cannot be evaluated in terms of averages. Nobel makes a further extension to an analogous law for weakly dependent (absolutely regular) observations.

3.5 The Relative Value of Labeled and Unlabeled Samples

In [Castelli and Cover], Vittorio Castelli presented new results on the relative value of labeled and unlabeled samples in reducing the probability of error of the classification of a sample into one of two classes based on the previous observation of labeled and unlabeled data. Castelli and Cover showed that when the training set contains an infinite number of unlabeled samples, the first labeled sample reduces the probability of error to within a factor of two of the Bayes risk. The Bayes risk is the best that a classifier could

do with complete a-priori knowledge of the densities associated with the two classes and the probability of each class. Further, they showed that subsequent labeled samples yield exponential convergence of the probability of error to the Bayes risk. Finally they show that labeled samples are exponentially more valuable than unlabeled samples with the relevant exponent being the Bhattacharya distance.

3.6 Feedback In Communication We report on progress in our efforts to find the capacity of the additive Gaussian noise channel with an average power constraint in the presence of noiseless feedback. In [Cover and Pombra], it was shown that the feedback capacity of a non-white additive Gaussian noise channel can be evaluated via the maximization of the determinant of a certain covariance matrix under a trace constraint. We have developed an algorithm for maximizing this determinant and conjecture that the values attained are indeed global maxima.

Our algorithm has given us insight into the optimum strategy for communicating with feedback in the presence of Gaussian noise. Preliminary results indicate that the majority of one's available power should be used to transmit linear combinations of past noise samples (which are available as a result of feedback) in an effort to reshape the effective noise spectrum. Specifically one should decrease the effective noise power where it is already low at the expense of increasing it where it is high.

3.7 Process Entropy

Various aspects of network information flow can be modeled using finite state Markov chains. Among these may be the lengths of message queues or the path a message follows as it is being routed through a network. We are therefore interested in calculating expectations of functionals of Markov trajectories. Laura Ekroot initiated this work in her PhD dissertation [Ekroot] and subsequent papers on the subject, [Ekroot and Cover a][Ekroot and Cover b].

Elza Erkip has generalized the problem of finding the entropy of Markov trajectories. Consider an irreducible finite state Markov chain with transition matrix P and gain matrix G , where the rs^{th} entry of G denotes the gain accumulated when the Markov chain passes from state r to state s . A trajectory T_{ij} is a path from state i to state j with no intervening states equal to j . We have found a closed form expression for the expected gain, g_{ij} for the trajectory T_{ij} .

As special cases of the general solution we obtain the expectations of various functions of trajectories which are expressible in terms of accumulated gains. Among these are the

length of the trajectory, the number of visits to state k on the way from i to j , the number of transits from r to s in the trajectory, and the entropy of a trajectory.

JSEP Supported Papers and Presentations

[Castelli and Cover] V. Castelli and T. Cover, The Relative Value of Labeled and Un-labeled Samples in Pattern Recognition, *1993 IEEE International Symposium on Information Theory*, San Antonio, Texas.

[Cover a] T. Cover, Complexity and Generalization in Neural Nets, *Neural Information Processing Systems Conference*, Denver, Colorado, December 1992.

[Cover b] T. Cover, What Processes Satisfy the Second Law? *IEEE Information Theory Workshop*, Salvador, Brazil, June 1992.

[Cover c] T. Cover, Results in Pattern Recognition and Neural Nets, *NEC Workshop*, Stanford, Calif., June 1992.

[Cover d] T. Cover, Duality of Data Compression and Investment, *Hewlett-Packard Seminar*, Palo Alto, Calif., April 1992.

[Roche b] J. Roche, Distributed Information Storage, Statistics Department Technical Report No. 79, Stanford University, March 1992.

[Ekroot and Cover a] L. Ekroot and T. Cover, The Entropy of a Randomly Stopped Sequence, *IEEE Trans. on Info. Theory*, 37(6): 1641-1644, November 1991.

[Ekroot and Cover b] L. Ekroot and T. Cover, The Entropy of Markov Trajectories, *Stanford Statistics Department Technical Report #78*, 1991.

[Pombra and Cover] Non-White Gaussian Multiple Access Channels with Feedback, submitted in 1991 to *IEEE Trans. on Info. Theory*,

JSEP Supported Ph.D. Theses

[Roche a] J. Roche, Distributed Information Storage, PhD Thesis, Stanford University, March 1992.

[Nobel] A. Nobel, On Uniform Laws of Averages, PhD Thesis, Stanford University, July 1992.

[Trott] M. Trott, Algebraic Structure of Trellis Codes, PhD Thesis, Stanford University, August, 1992.

[Pombra] S. Pombra, Non-white Gaussian Channels with Feedback, PhD Thesis, September, 1991.

[Ekroot] L. Ekroot, The Entropy of Markov Trajectories, PhD Thesis, September, 1991.

Other References

[Roche c] J. Roche, Multiple-User Distributed Information Storage, 1993 *IEEE International Symposium on Information Theory*, San Antonio, Texas.

[Cover and King] T. Cover and R. King, A Convergent Gambling Estimate of the Entropy of English, *IEEE Trans. on Info. Theory*, 24(4): 413-421, July, 1978.

[Willems] J. C. Willems, From Time Series to Linear Systems—Part I. Finite Dimensional Linear Time Invariant Systems, *Automatica*, 22(5):561-580, 1986

[Forney] G. David Forney, Jr., Geometrically Uniform Codes, *IEEE Trans. on Info. Theory*, 37(5):1241-1260, 1991.

[Loeliger and Mittelholzer] H. Loeliger and T. Mittelholzer, Convolutional Codes over Groups, submitted to *IEEE Trans. on Info. Theory*, 1992.